DOCUMENT RESUME

ED 335 410 TM 017 099

AUTHOR Chauvin, Sheila W.; And Others
TITLE Development and Validation of a Comprehensive Assessment System for Teaching and Learning.
PUB DATE Apr 91
NOTE 92p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 4-6, 1991). For related papers, see TM 017 059 and TM 017 062.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Beginning Teachers; Classroom Observation Techniques; Elementary School Teachers; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Learning Strategies; *Material Development; Secondary School Teachers; *State Programs; Teacher Attitudes; Teacher Certification; *Teacher Evaluation; Teaching Methods; *Validity
IDENTIFIERS *System for Teaching Learning Assessment Review; Teacher Performance Appraisal System; Validation Verification and Testing Techniques

ABSTRACT

This paper summarizes major research and development activities and reports the results of the development and validation of a comprehensive on-the-job statewide teacher assessment system designed to make inferences about enhancing student learning through classroom observation data--the System for Teaching and Learning Assessment and Review (STAR). The STAR was developed in response to legislative mandates in Louisiana; it is used to train principals, master teachers, supervisors, college faculty, and other educators to complete thorough assessments of beginning and experienced teachers' classroom performances for the purpose of renewable professional certification. Beginning in the 1988-89 school year and continuing through the 1989-90 fiscal year, work focused on the development, validation, and piloting of the STAR. The STAR represents a comprehensive dichotomous decision-making framework designed to assess key elements of effective teaching and learning. It consists of 117 assessment indicators that operationalize 22 teaching and learning components organized by four performance dimensions. The following analyses were conducted to establish the STAR's reliability and validity: validation studies; review of the research literature; content verification surveys; factor analyses; descriptive summaries of assessment data; external review; criterion-related validity studies; concurrent validity studies; reliability studies; standard-setting studies; external committee review; and qualitative studies. Data show that the STAR validly and reliably assesses effective teaching and makes inferences about student learning in a wide variety of classroom contexts. Nineteen data tables and a 46-item list of references are included. An appendix provides further details on the STAR, supplemented by 19 tables. (RLC)

Development and Validation of a Comprehensive Assessment System for Teaching and Learning

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
SHEILA W. CHAUVIN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Sheila W. Chauvin
Karen S. Loup
Chad D. Ellett

College of Education
Louisiana State University
Baton Rouge, LA

Paper presented at the annual meeting of the National Council of Measurement in Education, Chicago, Illinois

April, 1991

BEST COPY AVAILABLE

Development and Validation of a Comprehensive Assessment System for Teaching and Learning

Abstract

In response to legislative mandates, the state of Louisiana has supported the development of a comprehensive, on-the-job assessment system that is not only designed to assess effective teaching, but also to make inferences about student learning through classroom observation data. Unlike other large-scale "first generation" teacher evaluation systems, the STAR (System for Teaching and learning Assessment and Review) is clearly part of a new generation of assessment systems that "puts the light on the learner". Results of two years of extensive research and piloting offer convincing evidence that the STAR offers new horizons in the field of assessment of effective teaching and student learning.

Introduction

During the past decade, a variety of states have moved rapidly toward the development of on-the-job assessment/evaluation procedures for classroom teachers targeting certification, career ladder, merit pay, professional development and induction decisions. Beginning with the state of Georgia in 1980, approximately eighteen states have designed and implemented such assessment procedures and some twenty others are contemplating similar efforts. These large-scale efforts to assess teacher performance have been motivated by various accountability and educational reform policies established by state boards of education, state legislatures and some school districts. Indeed, "teacher assessment/evaluation" programs may very well be the cornerstone of current efforts toward educational reform. States such as North Carolina, Tennessee, Florida, Virginia, Kentucky, South Carolina, Arkansas, Missouri, Connecticut, New Mexico, Texas and others have followed Georgia's early lead to involve trained observers to complete relatively comprehensive evaluations of teachers (Chauvin, Ellett, Loup & Slan, 1990; Ellett, 1990).

In response to legislation provided in The Louisiana Teaching Internship Law (1984) and The Children First Act (1988), Louisiana has been involved in statewide efforts to develop a comprehensive, on-the-job assessment system to be used with all beginning teachers (1 to 2 years experience) in an internship program and in determining professional renewable certification of all 45,000 experienced teachers (3 or more years experience) in Louisiana. This system is called the

STAR (System for Teaching and learning Assessment and Review). The STAR has been designed to build on efforts of other states to identify and assess elements of teaching reflected in the extant process/product literature on effective teaching (Brophy, 1986; Porter and Brophy, 1986) and newer concerns about the assessment of knowledge of content, pedagogy and curriculum (Berliner, 1986; Shulman, 1986; 1987). The current version of the STAR (Ellett, Loup & Chauvin, 1990) also includes a variety of important assessment indicators new to the field such as indicators of the effective teaching of thinking skills and content structure and emphasis. Thus, the STAR is being developed in Louisiana in a way that moves the teacher assessment field forward in terms of "what" is measured within the context of a state mandate targeting the periodic, professional renewable certification of all teachers. In keeping with the statewide impetus for educational reform, the STAR assessment process is based on a model which: 1) puts the "light on the learner", 2) incorporates multiple assessors with multiple observations and 3) emphasizes on-going professional development based upon formative and summative assessment results.

Beginning in the 1988-1989 school year and continuing through the 1989-1990 fiscal year, two years of concerted efforts have focused on the development, validation and piloting of the STAR and corresponding processes for support and professional development for first-year teachers through the Teaching Internship Program (LTIP) and professional development, initial professional certification and continuing certification of experienced teachers through the Teacher Evaluation Program (LTEP). Throughout the research and development phases, and in keeping with state legislation, extensive efforts have been made to include input from and endorsement by classroom teachers and key educators (e.g., principals, assistant principals, instructional supervisors, college faculty and Department of Education personnel) in Louisiana. Statewide implementation of these programs (LTIP and LTEP) using the STAR began in October, 1990 with all first-year teachers (LTIP) and approximately 20% of all experienced teachers (LTEP). Current state legislation mandates that all 45,000 experienced teachers in Louisiana will have been assessed with the STAR by the end of the 1992-1993 school year.

Although statewide implementation has been initiated, research and development activities related to the STAR and these programs (LTIP and LTEP) are continuing and now include an additional focus on utilization of the STAR and these programs under real, "high stakes" conditions. In addition, current research and development efforts now target development of corresponding support and staff development programs, implementation studies and alternative applications of the STAR to other contexts (e.g., higher education).

Purpose

The purpose of this paper is to summarize major research and development activities and report results of the development and validation of a comprehensive on-the-job statewide teacher assessment system designed to evaluate teachers and make inferences about enhancing student learning through classroom observation data. While complete details of each developmental and validation activity are not provided in this summary paper, references are given for complete accounts of each investigative effort reviewed.

Instrument Development

Initial Development of the STAR

Legislative mandates forming the basis of the LTIP and LTEP require the development and implementation of a standardized, on-the-job assessment of teachers' classroom performances. Thus, the first project effort was to develop a draft assessment framework (instrument) and an assessor certification program to teach principals, master teachers, college faculty and other education professionals how to use the assessment framework according to a uniform set of assessment indicators and decision making rules. Many states have developed similar systems during the past ten years, and these systems served as an initial basis for the development of Louisiana's system. Given the requirements and intent of The Children First Act and the Louisiana Teaching Internship Law however, the system which has evolved in Louisiana extends these earlier teacher assessment efforts in other states in important ways. These include: 1) a more "student-oriented" focus in conducting assessments and in using assessment information to help teachers'

enhance students' learning; and 2) assessing a variety of important areas not given much emphasis in other states such as enhancing students' cognitive involvement in higher-order thinking and learning, and to stimulate effective use of thinking skills.

The initial development of the STAR began with a content synthesis of eight large-scale teacher evaluation instruments that had been designed in the late 1970's and early 1980's to assess on-the-job performances of teachers for a variety of purposes (Ellett, Garland & Logan, 1987). The eight instruments reviewed and synthesized for the initial development of the STAR assessment framework were the:

Teacher Performance Assessment Instruments (TPAI) (Georgia)

Georgia Teacher Evaluation Instrument/Process (GTEP) (Georgia)

Tennessee Career Ladder Teaching Evaluation System (TCLTES) (Tennessee)

Assessments of Performance in Teaching (APT) (South Carolina)

Virginia Teaching Proactices Record (VTPR) (Virginia)

Florida Performance Measurement System (FPMS) (Florida)

Teacher Assessment and Development System (TADS) (Dade County Public Schools) (Miami, Florida)

Texas Teacher Appraisal System (TTAS) (Texas)

These eight teacher evaluation instruments were believed to be the most thoroughly developed available and each was reasonably well-grounded in the extant research literature on teacher effectiveness. These instruments and their accompanying assessment processes had been designed to fulfill a variety of purposes such as providing support for beginning teachers, and to make teacher evaluation, certification and career ladder decisions. The content synthesis of these eight large-scale systems provided a strong research base for the initial foundation of the STAR, having been grounded in approximately fifteen years of prior research and development in other states.

Assessment items resulting from this content synthesis (n=620 individual descriptions/items) subsequently went through two content reviews by groups of Louisiana educators. The purpose of

these reviews was to identify and professional verify an initial set of assessment indicators to serve as a developmental framework for the STAR. This initial set of indicators was edited, structured and classified into various STAR Performance Dimensions and expanded to include more recent notions about effective teaching and learning. As the assessment framework of the STAR was developed, elements of the framework were further explicated by written Comments, Annotations and Decision Making Rules.

The first and most lengthy version of the STAR (151 assessment indicators) was piloted in Louisiana in 1988-1989. As a result of pilot research and development activities, the STAR was revised and a second, somewhat shorter version (140 assessment indicators) was piloted statewide in Louisiana in 1989-1990. As a result of this statewide pilot of the STAR, the STAR was reduced to a set of 117 assessment indicators now reflected in the current 1990-1991 version.

During the initial pilot of the LTIP and LTEP, two assessment instruments were developed .... one for beginning teachers and one for experienced teachers. The main difference between these two systems was the lesson planning requirements, with the LTIP requirement somewhat more thorough than the LTEP planning requirement. As a result of the 1988-1989 pilot and with input from educators from throughout Louisiana, the extended pilot version of the STAR was a unitary one, equally applicable to all teachers.

Many educators throughout Louisiana have contributed to the development of the STAR. As the STAR was piloted throughout Louisiana, input from teachers, principals, assistant principals, instructional supervisors, college faculty and other professional educators was incorporated into revisions of the STAR. Two years of statewide pilot activity included the involvement of more than 10,000 educators in Louisiana. During the 1989-1990 school year, for example, approximately 3500 principals, classroom teachers and other educators participated in a seven-day professional development program to be certified as STAR assessors. Many of these, in turn, shared information about the STAR and with teachers and others at their respective school sites. As a result, Louisiana educators have played a vital role in the development of the STAR as a comprehensive, "state-of-the-art" system designed to enhance the quality of teaching and learning in

a wide range of classroom contexts.

The current version of the STAR represents a comprehensive, dichotomous decision making framework designed to assess key elements of effective teaching and learning. It consists of 117 assessment indicators that require STAR assessors to use a set of common understandings and explicit decision making rules to make inferences about the quality and effectiveness of both teaching and learning. These decisions represent informed, professional judgments that must be made with careful consideration given to unique student, lesson and classroom context characteristics. The 117 assessment indicators operationalize 22 Teaching and Learning Components organized by four Performance Dimensions: 1) Preparation, Planning and Evaluation, 2) Classroom and Behavior Management, 3) Learning Environment, and 4) Enhancement of Learning. A copy of the overall organization of the STAR and a sample from the STAR manual are provided as Appendix A. The organizational framework shows the various Teaching and Learning Components for each STAR Performance Dimension and the number of assessment indicators defining each Component.

Assessment indicators and components comprising the first STAR Performance Dimension of Preparation, Planning and Evaluation are designed to make assessment decisions about the teacher's ability to plan for a five- to seven-day unit of teaching and learning. Emphasis is given to comprehensive planning in a manner that : 1) accommodates the range of students' needs, abilities and developmental levels; 2) structures the scope and sequence of content and curricula; 3) considers and specifies time allocations for teaching and learning activities; 4) considers and specifies appropriate materials, aids and activities that enhance student learning and the development of thinking skills; 5) carefully designs and specifies homework (Home Learning) assignments and formal assessment (student testing and evaluation) procedures.

The second STAR Performance Dimension (Classroom and Behavior Management) is operationalized by a set of assessment indicators and Teaching and Learning Components that reflect the teacher's ability to manage the total classroom learning environment including time, organizational and classroom routine tasks, student engagement in learning tasks, and acceptable

and unacceptable behavior. One component, Student Engagement, is not used to make certification decisions. However, it is an important assessment concern, given the well-demonstrated relationship between classroom engagement rates and subsequent student learning and achievement.

The third STAR Performance Dimension (Learning Environment) consists of two Teaching and Learning Components: 1) Psychosocial Learning Environment; and 2) Physical Learning Environment. Assessment indicators for these components reflect concern for a psychosocially supportive classroom climate and functionally effective learning environment built upon equity for students and positive interpersonal relationships between the teacher and students, and among students as well.

The fourth STAR Performance Dimension is termed Enhancement of Learning. This is the most lengthy Dimension of the STAR and it is comprised of nine Teaching and Learning Components defined by 55 assessment indicators. "Enhancement of Learning" implies that the teacher's role is one of a facilitator and guide for learning, rather than simply an "instructor", "trainer", or "deliverer of content." The assessment focus in this Performance Dimension is on the deliberate structure of "learning activities" in a way that allows students to be actively and cognitive engaged in learning and assume responsibility for their own learning.

Conceptual Basis and "Common Themes" of the STAR

The represents an effort to move the field of teacher evaluation forward by developing more comprehensive assessments of teaching and learning. This conceptual focus required grounding of the STAR in a variety of important "common themes". These common themes represent essential "key ideas" that permeate the philosophical basis and content of the STAR, the professional development program for certifying STAR assessors, the STAR assessment process, assessment decisions about the quality and effectiveness of teaching and learning, and corresponding professional development modules and resource materials. These themes are thoroughly discussed elsewhere (Ellett, 1990) and will not be detailed here. However, a list of

these common themes are as follows:

- All students can learn
- Teaching and learning
- Teaching/Learning as a total process
- Learning to learn/Self responsibility for learning
- Role of preparation, planning and evaluation (reflective practice)
- Knowledge of...
  a. Pedagogy
  b. Content
  c. Curriculum
- Time
- Active involvement/engagement
- Individual differences
- Quality learning environment
- Cognitive development/thinking skills

Each of the STAR common themes listed above represents a conceptual "thread" that ties the STAR content and assessment process together as a holistic, contextually-based assessment system. Unlike many simpler teacher evaluation instruments and checklists, the STAR content and assessment processes have been deliberately designed to define effective teaching practices in terms of their linkages to student interest and involvement in learning. Also, the focus of assessment with the STAR is not only teacher behavior, but the wide variety of teacher-student and student-student interactions. As a contextually-based assessment system, the STAR also requires assessors to make assessment decisions about the quality and effectiveness of teaching and learning by carefully considering the unique context characteristics of each classroom. A complete discussion of the structure and decision making framework in the STAR, as it is designed as part of a new generation of teacher assessment systems and "puts the light on the learner", can be found in Ellett (1990).

STAR Assessor Certification Program

A key component of any large-scale assessment system is the program designed to prepare and certify educators to use the assessment framework. During the 1988-1989 a comprehensive, 8 1/2 day model was developed and piloted with approximately 375 educators (principals, master teachers, instructional supervisors, college faculty and other key educators) in six regions of Louisiana. A total of fifteen, sessions were completed during the spring of 1989. Participants represented virtually every school district in Louisiana. Input and suggestions were obtained each day of every session, resulting in revisions and modifications in the program, program materials and the STAR literally after each regional session.

In addition, a three-day STAR "program assistant" certification program was developed and piloted with approximately 90 educators certified in the pilot STAR assessor program and recommended for this additional role. Likewise, input was sought and revisions were made in the program, materials and STAR.

A comprehensive review of programs and materials was conducted in June 1990, with a panel of pilot-certified educators from across Louisiana. Revisions and modifications were completed and two "field tests" of these revisions were conducted during the summer of 1989. Additional revisions were made before the beginning of the 1989-1990 extended pilot year. The assessor certification program was refined and shortened to seven days and the "program assistant" certification model was shortened to two days.

During the 1989-1990 extended pilot, STAR assessor certification programs were conducted in twelve sites from October through May. At two week intervals, twelve new sessions were initiated with thirty participants per session (mixed by parish and position types) in ten regions of Louisiana. Thus, every two weeks a new group of 360 educators entered STAR assessor certification programs. Sessions were conducted by certified program leaders, who were prepared, certified and supervised by project staff at LSU. Approximately 120 STAR assessor certification programs were conducted during the 1989-1990 extended pilot year, with approximately ten STAR program assistant sessions completed regionally, as well. Throughout the

extended pilot year, ongoing assessment of program activities and multiple proficiency requirements for certification provided input and suggestions for resulting revisions in each program and accompanying materials.

Research

The research agenda to support the psychometric quality of the STAR and assessment process was begun during the spring of 1989. A variety of data were collected to establish the validity and reliability of the STAR and to examine the quality of teacher performance with in-field assessments of actual classroom teaching. A brief summary of each of these research studies is provided. More detailed explanations of these research studies, their results and implications are available in a series of technical reports. These are referenced accordingly throughout this paper.

STAR Validation Studies

A variety of research and development studies bearing on the validity of the STAR was completed during the 1988-1989 and 1989-1990 pilot years. Research and development studies continue during the current year (1990-1991) as the STAR is being used in statewide implementation of the LTIP and LTEP. A brief summary of each of these studies is provided in the following sections.

Use of the Research Literature in Teaching and Learning

As the STAR was developed during the first pilot year, pertinent research and theory-based literature on effective teaching and learning was reviewed. Results of this review were aggregated and reported to document assessment indicators and components of the STAR relative to existing research and to "ground" the STAR in past attempts to link important elements of teaching to student outcomes (Claudet & Ellett, March, 1990). This is an important and ongoing effort in establishing the construct validity of the STAR. As the research literature and theory base for effective teaching and learning continue to develop, the document providing review of the literature pertinent to the STAR is continually updated (Claudet & Ellett, 1990). Though the STAR content reflects important elements of the research base on effective teaching and learning, and it is a "research-based" assessment framework, one is cautioned against over-extending the extant research

documentation in making this claim. Thus, two kinds of support for the "research base" of the STAR need to be considered as it continues to develop: the extant literature in effective teaching and learning, and actual research with the STAR in Louisiana classrooms. Both of these continue to be ongoing efforts, even as these programs are currently being implemented statewide under real, "high stakes" conditions. The initial, selective review of the literature on effective teaching and learning provides support for the validity of the STAR as a system reasonably well grounded in this literature.

STAR Content Verification Survey

1988-1989: During the late spring of 1989, a random sample of 6,000 teachers representing every school district in Louisiana was selected for a survey to professionally verify an initial set of teaching and learning components of the STAR. The survey form requested that participants make several professional judgments about each STAR component. These judgments, stated in the form of more simple questions were as follows: is the particular STAR Teaching and Learning Component 1) clearly stated? 2) applicable to the subject you teach? 3) free of bias? 4) a reasonable performance expectation? and 5) essential to the enhancement of student learning? The survey also requested that participants indicate the degree to which they believed beginning and experienced teachers were prepared to demonstrate performance in the various performance dimensions comprising the STAR.

Useable results were received from approximately 2300 teachers from throughout Louisiana (response rate = 38.3%). By way of summary, the results showed strong endorsement from Louisiana teachers of the basic elements comprising the STAR for the questions asked. Percentages of endorsement for the various questions asked typically exceed 90% of the teachers responding. Overall, 92% of the 2300 respondents supported the STAR Teaching and Learning Components as reasonable expectations for teachers seeking initial, professional certification in Louisiana, and 89% for teachers seeking renewal of professional certification in Louisiana.

When considering the degree to which beginning and experienced teachers are prepared to successfully demonstrate performance in teaching and learning components of the STAR, 40-55%

of the respondents indicated that experienced teachers were completely prepared, compared to only 14% of the beginning teachers. The results indicated that 28% of the respondents believe that beginning teachers are "not prepared at all" to successfully meet expectations in the STAR dimension of Classroom and Behavior Management. A complete report of this content verification effort is provided in Ellett, Naik & Logan (1990).

1989-1990: A second content verification study was conducted during the late spring of 1990. A similar survey to the one used in the initial pilot year was used to survey approximately 400 "expert" educators to verify STAR assessment indicators as being reasonable expectations for beginning and/or experienced teachers in Louisiana. The focus in this survey was on STAR indicators, since these represent the fundamental decision making level in the assessment process. Thus, this study sought to verify STAR content at the assessment indicator level using perspectives of Louisiana teachers and other informed educators, so that educators from throughout the state, representing all school districts and teaching and learning contexts, would have input into an important assessment and support process. Of the approximately 400 educators in the sample for this content verification study, 60% were classroom teachers, 30% were school administrators, and 10% were instructional supervisors and college faculty. All participants in the sample were nominated by STAR program leaders as having an "expert" understanding of STAR assessment indicators. Each individual had successfully completed from seven to twelve days of intensive professional development and were considered highly knowledgeable and experienced with the draft STAR and assessment procedures. Methodology and data collection and analyses procedures were the same as those used in the initial content verification study. 336 useable surveys were returned, yielding a response rate of 84%.

The content verification of each set of indicators comprising each teaching and learning component indicated that an overwhelming majority of respondents endorsed each of the indicators as applicable to their subject area or content specialty. In addition, the results strongly supported most of the indicators as reasonable performance expectations for both beginning and experienced teachers and, in most instances, greater support for the indicators as important to the enhancement

of student learning. Considered collectively, the results show strong teacher endorsement of the appropriateness and applicability of the indicators and teaching and learning components to elementary and secondary settings. The data also supported the indicators and components in terms of observability and freedom from bias against any particular group of teachers (e.g., gender, ethnicity, etc.).

Somewhat less support was evidenced for specification of time for each major teaching and learning activity (STAR Performance Dimension I, Component C: Allocated Time and Content Coverage) as being applicable to their subject area or content specialty, as a reasonable performance expectation and as important to the enhancement of student learning than for other STAR components.

Stronger endorsement of the Comprehensive Unit Plan (CUP, Performance Dimension I) was evidenced for initial certification than for renewing certification. Overall, results indicated a range of approximately 5-15% experienced teachers and 10-25% beginning teachers may not be adequately prepared to successfully address STAR components. However, respondents indicated that statewide orientation should reduce these percentages.

Results provided in this study verify the "job-relatedness" of assessment content. Also, results include professional "expert" judgments about many new criteria (e.g., indicators addressing teaching thinking skills) on the STAR not well represented on other state assessment systems. A complete description of this study, results and implications may be found in Ellett, Chauvin, Loup & Naik (1990).

Factor Analyses

Two series of factor analyses have been conducted to explore and verify/confirm the construct validity of the STAR as a comprehensive, classroom-based assessment system of teaching and learning.

1988-1989: An initial series of factor analyses was conducted to confirm the logical classification of assessment indicators via a series of STAR Teaching and Learning Components. The sample for the study consisted of 933 classroom teachers drawn from public schools

throughout Louisiana. These teachers were randomly selected "volunteers" that were asked to participate in STAR assessments by principals, master teachers and supervisors to meet field assessment requirements of a program to certify these educators as STAR assessors. These teachers were randomly selected within participating schools from alphabetical faculty lists submitted by STAR assessors. The teachers represented a majority of classroom contexts found in the public schools including special education, music, art, vocational settings and so on.

Data for the study were collected by a group of approximately 350 principals, master teachers, supervisors and other educators participating in the statewide pilot of the STAR, LTIP and LTEP. The 1989 version of the STAR used in this study consisted of 151 assessment indicators that operationalized 23 teaching and learning components organized by four performance dimensions. Two kinds of data analyses were completed in this study. First, a summary of descriptive statistics for each STAR assessment indicator was made. This summary provided information about "mastery" levels relative to each assessment indicator. Secondly, a series of oblique and orthogonal factor analyses of STAR assessment indicator scores was completed as an initial "probe" as to the extent to which indicators and components seemed to "hang together" as they were originally classified.

The results of the factor analyses of classroom observation data collected with the STAR provided some useful information about the construct validity of the STAR as a comprehensive measure of teaching and learning. Interestingly, 87 of the 117 STAR assessment indicators (Performance Dimensions II, III and IV) maintained their original classifications by the various teaching and learning components. This finding tends to support the logical classification of the STAR assessment indicators when the content of the STAR was originally constructed. Assessment indicators in components such as "Psychosocial Learning Environment," "Sequence and Pace," and "Content Accuracy and Emphasis," seem to be more "spread out" across factors. However, this seemed logical since affect, order, pace and clarity seem to pervade teacher and student behaviors throughout a lesson.

It is important to note that analyses were conducted on single assessments of teaching and learning across the teacher sample. Results of these initial analyses were useful in "fine tuning" the STAR before being used in actual implementation of the LTIP and LTEP. Also, results were used to develop a computerized summary profile for use by teachers in developing continuing professional development plans. Ellett, Loup, Chauvin & Naik (1990) provides a complete description of this study, its results and conclusions.

1989-1990: A second series of factor analyses were conducted in the spring 1990 to further confirm the classification of STAR assessment indicators derived from the process/product and human learning literature. Also, some minor reclassification and revision in the original organization of the STAR had been completed based on results of earlier factor analyses (as described above). These factor analyses were also an attempt to confirm or verify current classification of assessment indicators within existing teaching and learning components using a larger sample.

The sample for this study was the classroom performance of teachers and their students in 5,720 classrooms derived from a random sample drawn from all 66 public school districts in Louisiana. Both teachers and their students were included in the sample, since the STAR assessments required trained assessors to score assessment indicators giving consideration to teacher behaviors, teacher-student interactions, student-student interactions, student engagement rates and student active involvement, interest and participation in learning tasks.

Data were collected in actual teaching and learning settings using the 1989-1990 extended pilot version of the STAR. Assessors using the STAR had been certified through a comprehensive, seven-day professional development program. Assessors included principals, assistant principals, master teachers, instructional supervisors, college faculty and other professional educators who had successfully completed certification requirements as a STAR assessor. Descriptive statistical summaries for assessment indicators and teaching and learning components were completed. Also, results of a series of factor analyses were completed using SAS PROMAX procedures conducted in an iterative fashion to examine the original classification of the assessment indicators by each

teaching and learning component.

The initial factor analysis was a one-factor solution. Approximately 22% of a total of 117 assessment indicators did not significantly load on a single factor (loading less than .33). This one-factor solution also accounted for only 21% of the total variation in the data. A series of orthogonal analyses were completed in an iterative fashion extracting two to twenty factors. Examination of these various solutions suggested that a sixteen-factor solution best fit the original classification of STAR indicators via the various teaching and learning components. This solution accounted for approximately 52.4% of the total variation in the data. Factor loadings (factor/indicator correlations) ranged in magnitude from approximately .33 to .93, with .60 being most typical. Interestingly, all 117 indicators significantly loaded (at least .33) on one or more factors. Twenty indicators loaded on more than one factor. However, for the most part, the patterning of loading confirmed the original classification of STAR assessment indicators by the various teaching and learning components. In some instances, such as "Psychosocial Learning Environment", assessment indicators loaded on more than one factor. However, this is consistent with the view that affective elements of the learning environment, for example, are pervasive throughout other aspects of teaching and learning interactions. Newer assessment components reflected on the STAR, for example, the teaching of thinking skills, were confirmed by these analyses as independent factors. Again, as with the first series of factor analyses, data were collected as single assessments. A complete description of this second series of factor analyses may be found in Ellett, Loup, Chauvin, & Naik (1990).

Both series of factor analyses generally support the original classification of STAR assessment indicators and provide convincing evidence that these indicators of effective teaching and learning are factorially independent and assessors can be taught to differentiate these indicators without being over stringent or generous. Future factor analytic studies will focus on analyses of data collected in a manner consistent with the assessment process (i.e., multiple assessors over multiple occasions) and under real, "high stakes" conditions.

Descriptive Summaries of STAR Assessment Data

As part of the initial pilot of the STAR, field data were collected from 969 assessments of teaching and learning in classrooms, in virtually every school district in Louisiana. Data were collected by some 350 principals, master teachers and other Louisiana educators who were pilot-certified STAR assessors. No Comprehensive Unit Plans (CUPs) were assessed and only single observations of lessons occurred. No teacher was assessed more than once, and various assessors completed from two to five assessments. Teachers were randomly sampled from alphabetical faculty lists provided by educators participating in the pilot STAR assessor certification program during the spring 1989. Thus, data collected represented the wide variety of contexts in which teachers work and are now currently being assessed to meet the requirements of new Louisiana laws.

Data analyses were completed to examine the various levels of teacher performance relative to the STAR assessment indicators, teaching and learning components and performance dimensions. Statistical summaries of these data were made for STAR teaching and learning components and assessment indicators in terms of the percentage of "acceptable" and "unacceptable" assessment decisions. Table 1 presents a summary of the percentage of maximum possible scores for each STAR teaching and learning component. These results indicate the percentage of "acceptable" decisions made by STAR assessors for the total number of assessment indicators comprising each Teaching and Learning component summed over all 969 assessments completed. For example, for the teaching and learning component of TIME in "Classroom and Behavior Management", 8 assessment indicators X 969 assessments generates a maximum of 7752 decisions. The last column in Table 1 shows the percent of the maximum possible score to be 73.41%. The lowest performance area reflected in these assessments was in "Thinking Skills", and high areas of performance were in "Physical Learning Environment" and "Oral and Written Communication." More typically, 20-40% of the assessment decisions for the various STAR indicators were "unacceptable".

It is important to note that these data were collected by Louisiana educators typically in their own school or district. Also, teachers assessed with the STAR had little orientation to the STAR content and assessment process. Ellett, Chauvin, Loup & Naik (1990) provides a complete report of this study.

1989-1990: A second analysis of elements of effective teaching and learning derived from 5720 single classroom-based assessments with the STAR was conducted to further define a reliable data base to make inferences about the effectiveness of everyday teaching practices and to compare the effectiveness of teaching and learning by school level and major subject area.

Using the 1989-1990 extended pilot version of the STAR, approximately 3000 educators (i.e., principals, master teachers, instructional supervisors, college faculty and other educators) pilot-certified as STAR assessors collected data in virtually every school in every district in Louisiana. Methodology and data collection procedures were the same as those used in the initial pilot study. This large sampling of Louisiana teachers encompassed a wide variety of both subject areas and teaching and learning contexts, and thus reflects the kinds of assessment situations in which teachers will be observed/assessed for the purposes of induction and renewable certification.

In addition to summarizing results by percentage of "acceptable" and "unacceptable" decisions by component and indicator levels for the total sample, "between-groups" comparisons were also made. Comparisons were made for elementary versus secondary contexts, beginning versus experienced teachers, and "cognitive-based" versus "performance-based" classrooms. Table 2 provides a summary of the percentages of the maximum possible scores for each teaching and learning component for the total sample. Detailed descriptions and results of each analysis in the study may be found in Claudet, Hill, Ellett & Naik (1990).

The results of the descriptive and comparative analyses provided some interesting insights into everyday practice and "life" in classrooms. Overall, the results indicated that less than 50% of the total possible assessment decisions for the sample of 5720 classrooms observed were assessed as "acceptable" in areas such as student engagement, managing task-related behavior, lesson and activities initiation, content accuracy and emphasis, monitoring learning tasks and informal

assessment, and feedback. Only 22% were assessed as acceptable in developing students' higher order thinking skills. STAR teaching and learning components with acceptable decisions at or above 75% of the maximum possible scores included oral and written communication and the physical learning environment.

Results from both descriptive and comparative analyses (1988-1989 and 1989-1990) have shown few major differences at the STAR component level between elementary and secondary classroom settings and between beginning and experienced teachers. The greatest differences noted between beginning and experienced teachers seems to be in the area of "Classroom and Behavior Management", favoring experienced teachers in "acceptable" decisions. Of greatest concern and as evidenced in both studies was the overall low performance levels in structuring and involving students in learning tasks that enhance the development of thinking skills.

External Review of the STAR

As part of the construct validation process, and particularly in an effort to support the content validity of the STAR, an external consultant was used to select external "expert" consultants to review and critique the STAR in terms of content, clarity and measurement application to elements of effective teaching and learning across the full range of classroom contexts. Results of these external "expert" reviews provided much evidence in support of the content validity of the STAR as an assessment/measurement system for effective teaching and learning. Suggestions for enhancing the quality of the STAR offered by these external reviews were incorporated into revisions made in the STAR near the end of the 1989-1990 extended pilot year and are reflected in the current 1990-1991 version of the STAR. A summary of these external reviews may be found in Tobin (1990a).

Criterion-Related Validity

Criterion-related validity studies of the STAR assessment framework and process have been completed during FY 1988-1989 and 1989-1990. These validation efforts were designed to probe the extent to which relationships could be established between assessments of the quality of teaching and learning using the STAR and three important, student-related criterion variables: 1)

student achievement on teacher-made tests; 2) student perceptions of elements of the classroom learning environment; and 3) classroom indices of active engagement in learning. These three variables were selected as part of the validation effort because of their importance as predictors of learning and subsequent student achievement and their implications for overall construct validation of the STAR.

In the initial criterion-related validity study, conducted during the late spring 1989, data were collected from a sample of 66 classrooms (30 elementary/grades 2-6 and 36 secondary/grades 7-12) using STAR assessment teams (i.e., principal, master teacher, "outside assessor"). These classrooms were selected from a larger sample of schools in the district giving consideration to a reasonable balance among school size, socioeconomic status (SES), attendance (ADA) and other characteristics so as to reflect demographics of the total district. Each team modeled the STAR assessment process of independent observations on each of two occasions, resulting in six STAR assessments for each teacher. Teacher-made test data and student perceptions data were collected from all students in each teacher's class over a 7 to 10 day unit of teaching and learning. These data were processed and analyzed using class means as units of statistical analysis.

Two kinds of analyses were computed in the study: 1) descriptive statistics for elementary, secondary and total classroom groups; and 2) Pearson Product-Moment correlations among various variables for elementary, secondary and total classroom groups using class means as the units of analysis.

A variety of interesting findings emerged from these analyses that bear on the criterion-related validity of the STAR. For example, strong positive relationships (correlations) were established between class engagement rates and the quality of teacher performance as assessed by the STAR. This finding is highly encouraging since class engagement rates have repeatedly been shown to be a strong correlate, in turn, of long-term student achievement gains. Positive relationships were also evident (particularly for elementary classrooms) between student perceptions of important characteristics of the classroom learning environment and teacher performance as assessed by the STAR. Also of note, the STAR Teaching and Learning Component of Thinking

Skills was positively and significantly related to achievement gain in both elementary and secondary classrooms.

There was little relationship between the student engagement rate index and achievement gain. Thus, quantitative indices of student engagement in learning tasks may not be sufficient to enhance meaningful learning and subsequent achievement. This finding suggested that even though the "quantity" of engagement may be quite high, the overall "quality" and "intensity" of engagement may be rather low. As a result, this line of inquiry was pursued in a subsequent study conducted during the extended pilot year (1989-1990). Tables 3-6 provide summary data of these results. A complete description of the study, results and implications may be found in Ellett, Loup, Chauvin & Naik (1990).

1989-1990: A follow-up investigation, conducted during the 1989-1990 extended pilot year, used a broader sample of Louisiana classrooms, more experienced STAR assessors, over longer periods of time (approximately six weeks) than in the prior initial research efforts. The sample for this second study consisted of teachers and all students in 66 classrooms selected from two large, urban school districts in Louisiana. The classrooms were selected from a larger sample of schools in the districts giving consideration to a reasonable balance among organizational patterns (elementary, middle, high school), subject matter taught, teacher experience (student teacher, beginning teacher and experienced teacher), socioeconomic status (SES), and other characteristics, so as to reasonably reflect demographics of the district.

In this study, 40% of the teachers were asked to prepare a Comprehensive Unit Plan (CUP) for the first set of assessments. STAR Performance Dimensions I, II, III and IV were assessed. Other participating teachers provided STAR assessors with a daily lesson plan and information about classroom and student characteristics to assist with framing the context for subsequent classroom observations.

Methodology, data collection and analyses for this study were similar to the research design used in the initial investigation. A different paper and pencil measure of students' perceptions of the learning environment was used for secondary students (Classroom Learning Environment Scale

[CLES]). Also, the quality and intensity of engagement was also determined. As engagement scans were made on each of the STAR assessment occasions, each assessor recorded for those students engaged, the percentage who were engaged at high, middle and low levels of quality and intensity. Tables 7-13 provide summary results of this study.

The relationships between assessments of STAR teaching and learning components and student perceptions of the learning environment remain less frequent and not as strong as desired, but analysis of specific relationships provided additional insights. These are discussed in the complete report (Lofton, Ellett, Chauvin, Loup & Claudet, 1990).

The findings for the achievement gain index using teacher-made tests are of particular interest because they suggest that future validation research with the STAR has the strong potential to demonstrate positive and significant relationships between STAR performance levels and student learning. Of concern is the general quality of teacher-made tests for future studies. Many of the tests developed for this study showed pretest "ceiling" effects. As these tests become more reliable, validity evidence for the STAR should be more frequently occurring and even stronger than that obtained in this study.

In summary, the results obtained in this study and reported in Lofton, Ellett, Chauvin, Loup & Claudet (1990) are encouraging and continue to support the criterion-related validity of the STAR in Louisiana's classrooms. The correlation coefficients are within the range of typical criterion-related validity coefficients for other measures of teacher performance and many exceed this range. For example, in a review of the process/product literature, Medley (1977) suggested that correlations of .30 to .40 between classroom-based teacher observation/evaluation measures and student outcomes are sufficiently strong to support criterion-related validity....though they are not evident in very many studies using indices of student achievement as a criterion variable. Many correlations in this study exceed this range in magnitude and many are higher than those reported in Medley (1977) and other more recent research syntheses. Also, results obtained in this study are very encouraging, given the rather small sample sizes. A larger sample would have contributed to more statistically significant results.

The index of student engagement in learning tasks continues to show the greatest validity with the STAR. The results of this study also provided some new insights into the importance of examining indices of the "quality and intensity" of student engagement in learning tasks in future STAR validation efforts. The engagement correlations suggest that teachers who score high on the STAR maintain student engagement at high rates with high quality. Those scoring low on the STAR are teaching in classrooms with higher percentages of students engaged in learning tasks with low quality and intensity. The linkages established here between STAR performance levels and student engagement in learning seem to support the "ecological" validity of the STAR as a measure of both effective teaching and learning.

Concurrent Validity

A study was conducted in spring 1990 to examine the extent to which the STAR can differentiate "superior" teachers from other teachers. This is of particular concern, since legislation underwriting the LTEP requires identification of "superior" performance levels on the STAR as one of the qualifications for entering a career option program (Model Career Options Program/MCOP). This study also provided an opportunity to examine the validity of teacher's holistic, high inference judgments about their colleagues, and to examine and compare actual classroom performances associated with these judgments.

The sample consisted of 100 teachers from public schools throughout Louisiana, balanced by grade level and SES. Regional LTIP/LTEP coordinators recommended schools where both the principal and master teacher had successfully completed STAR assessor certification requirements and would volunteer to participate in the study. All teachers in these 100 schools were asked to confidentially nominate at least one and no more than three excellent teachers on their faculty who "routinely perform in the classroom at only the most outstanding levels of excellence and in a manner that consistently enhances student learning". Approximately 2300 nominations were received from teachers in these 100 schools. Proportions of nominations were computed for those nominated and teachers were ranked according to the percent of nominations received. From this ranking of nominated teachers, the 50 teachers with the highest percent of nominations were

identified as the known group of "superior" teachers in the study. Since teachers at each school were asked to nominate three teachers, the highest possible percent of nominations a teacher could receive was 33%. The percent of nominations received by this group of teachers ranged from 14% to 31%. In each of the 50 schools, a randomly selected teacher was chosen from the remaining teachers. The percent of nominations received by this group ranged from 0% to 15%. Additionally, a random "comparison" sample of 26 teachers who received no nominations was selected from faculty lists.

The 1989-1990 extended pilot version of the STAR was used to collect data during the spring 1990 from classrooms throughout Louisiana. Teachers were asked to voluntarily participate. Each teacher in the study was assessed by a three-member team which included the principal and master teacher from the teacher's school and an outside assessor. All assessors were certified in the use of the STAR.

Assessors were not told whether the teachers they observed belonged to the "superior", "random", or "comparison groups. Each member of the team assessed the teacher on two occasions, with a minimum of ten school days between the first and second observation for each assessor. All observations were announced visits, and teachers chose the subject and class periods during which observations would take place. The Comprehensive Unit Plan (CUP) was not a part of the assessment. Only classroom observation data needed to make assessment decisions about STAR assessment indicators in Performance Dimensions II, III and IV were collected. To assist assessors, teachers were asked to supply a copy of their regular daily lesson plans. After six observations, teachers could request a copy of their assessment profiles; however, no feedback was provided by individual assessors. Complete data sets for 87 teachers were obtained and used in analyses.

Data from STAR assessments were aggregated across six assessments for each teaching and learning component by teacher groups. Analyses of descriptive data were completed. Mean numbers of acceptable decisions and percentages of the maximum possible scores ("mastery" scores) for each STAR component were computed for each group. In addition, mastery scores

were also compared to a set of "benchmark" standards, a recommended percentage of acceptable decisions for each component to be piloted during the first year of implementation of the STAR program for professional renewable certification, to determine success rates for each group.

A series of one-way analysis of variance procedures were used in comparing the three groups of teachers using component score means. Scheffe's post hoc comparison technique was used to locate significant ($p<.05$) differences. An important aspect of these analyses was the extent to which teachers in the "superior" group scored differently than those in the "random" and "comparison" groups.

The results of group comparisons provided some interesting findings. First, as shown in Table 14, "mastery" scores for the superior group exceed those for both the random and comparison groups on all 16 STAR Teaching and Learning Components (excluding II. C., Student Engagement). Mastery scores for teachers in the random group were consistently higher than scores for those in the comparison group with the exception of Teaching and Learning Component II. C., Student Engagement, where mastery scores were equivalent. The overall implications of these results indicate that the STAR process can differentiate teachers across assessors and occasions for Teaching and Learning Components.

The results of analysis of variance comparisons of STAR component mean scores, provided in Table 15, showed significance ($p<.05$) favoring higher scores for the superior group, when compared to the random or comparison groups on 14 of the 17 components. However, post hoc comparisons revealed significant differences ($p<.05$) between superior and random teachers for only one component, Thinking Skills. Significant differences were noted between the superior and comparison groups on all 14 of the aforementioned components, as well as between the random and comparison groups on 6 of the 14 components.

Examination of the distribution of scores within the three groups showed considerable overlap with some teachers, particularly in the superior and random groups, indicating that some teachers in the random group may have received a portion of peer nominations and some teachers who received no nominations may have actually received higher component scores than many of

those who were designated "superior". Thus, using only teacher nomination criteria for differentiating teachers along a continuum of effectiveness of teaching and learning appears only partially accurate and somewhat unreliable. However, the large differences in superior and comparison teacher groups point to the fact that the STAR can clearly discriminate superior teachers.

Another analysis was conducted with data collected in this study. Using a set of "benchmark" standards recommended by a committee of Louisiana educators, primarily classroom teachers, a comparison of each group (superior, random and comparison) was made regarding the predicted percentage performing below these recommended expectations for each STAR component. Table 16 provides a summary of these results. While similar results were obtained in a number of teaching and learning components for superior and random teachers, larger differences were clearly evident when superior teachers were compared to comparison teachers. The only obvious exception was in the teaching and learning component of "Feedback". Random teachers seemed to "outperform" both superior and comparison teachers, but all three groups were very similar in performance levels. A complete description of this study and discussion of results and conclusions may be found in Ellett, Loup, Chauvin, Lofton & Naik (1990).

STAR Reliability Studies

Investigations of the consistency and stability of data collected with the STAR were conducted during the 1988-1989 and 1989-1990 pilot years. Similar studies are continuing as statewide implementation has been initiated and assessments are now completed under real, "high stakes" conditions. Results reported here represent findings obtained under pilot and research conditions. Two kinds of reliability analyses have been completed as part of STAR research activities during the two-year pilot program. Internal consistency reliabilities were computed for STAR performance dimensions and teaching and learning components. Results of analyses completed during the first pilot year (1988-1989) showed reliabilities within an acceptable range (.75 to.98). Secondly, two "generalizability" studies have been completed to assess the extent to which the STAR assessment framework and process (three-member team on two occasions) could

adequately differentiate teacher performance and generalize assessment results over STAR assessment indicators and assessment occasions.

The reliability model used reflects a comprehensive data collection system similar to those developed in the past in other states such as Georgia. Past investigations of the reliability of these systems that include the use of multiple data collectors over multiple occasions have proven to be quite promising (Capie, Tobin, Ellett & Johnson, 1981; Capie & Ellett, 1982; Performance Assessment Systems, 1984). Reliability studies of the STAR summarized here extends this work, since the STAR has been designed to assess the effectiveness of teacher performance and student learning at the same time.

All analyses were completed using A General Purpose Analysis of Variance System (GENOVA) (Crick & Brennan, 1983). Generalizability theory (Brennan, 1978; Crocker & Algina, 1986; Cronbach, Gleser, Nanda & Rajaratnan, 1972; Medley & Mitzel, 1963) was selected as the method of choice for the analyses. In its derivation from analysis of variance, GENOVA allows for identifying and estimating multiple sources of variation simultaneously. Also, it has the added benefit of providing for the simulation of alternative data collection strategies such as variations in numbers of observers or observation categories. A properly designed study which generates a high generalizability coefficient provides evidence that the assessment system can differentiate subjects (i.e., teachers) in terms of their abilities, while generalizing over assessors (i.e., agreement among principal, master teacher and outside assessor), items (i.e., internal consistency of assessment indicators and components) and assessment occasion (i.e., stability from fall to spring assessments). When coefficients are lower than desired, examination of variance components for facets in the design can suggest where there may be undesirable variation in the data.

1988-1989: An initial generalizability study was conducted during the late spring 1989 in eleven schools in an urban school district in southeast Louisiana. Altogether 46 teachers were assessed on the STAR on two occasions by each of three observer types (principal, master teacher, outside assessor). All data were collected confidentially, and no discussion of results with assessed teachers occurred until all six observations were completed and summarized. A total of 276

assessments were completed (46 teachers X 6 observations).

The observers in this study were trained by project staff immediately preceding data collection. All assessors, except for outside assessors, completed an abbreviated 4-5 day preparation program and considered to be proficient enough to conduct accurate assessments. This was a limit of the study and noted in the full interpretation of results. The outside observers were project staff members, who had not only been prepared and certified in the use of the STAR, but had also been extensively involved in teaching and certifying other educators as STAR assessors.

Data collection procedures yielded scores for seventeen components across three performance dimensions: STAR Performance Dimensions II, III and IV. Performance Dimension I was not analyzed in this study, as teachers were not asked to complete Comprehensive Unit Plans (CUPs).

The data from this initial generalizability study provided a preliminary estimate of the reliability of the STAR as a data collection system. A summary generalizability coefficients for each teaching and learning component is provided in Table 17. The average generalizability coefficient with the effect of all three assessors considered was .67. Results of initial generalizability analyses showed coefficients in the range of .45 to .73 for a two-person team (principal and outside assessor) and from .50 to .81 for a three-person team (adding the master teacher). Given the preliminary nature of this study, a generalizability coefficient of this magnitude seemed reasonable, and is consistent with those for other on-the-job assessment systems reported elsewhere (Capie, Ellett & Cronin, 1985).

In general, there seemed to be consistent decisions among the three assessor types across the 17 components assessed by the STAR. In fact, if the percentage scores given by the three assessors were correlated, the following results were found: the correlation between the principals' percentage scores and the outside assessors' percentage scores across the 17 components was .97; the correlation between the principals' and the master teachers' scores was .91; and that between the master teachers' and outside assessors' scores was .95. Thus, results obtained suggested that as percentage scores by components increased for one group of assessors, they also increased for the

other assessor groups. Based on initial results, average percentage scores across components in the STAR appeared highly consistent. For example, all three assessors judged teachers highly on "Physical Learning Environment", while judging them relative low on "Thinking Skills". These results suggest common perspectives across assessor types, as they view classroom teaching and learning over multiple teachers and multiple lessons (occasions).

While all three assessor types agreed in terms of the relative percentage of teachers satisfactorily mastering components, there were some differences in their assessments. Master teachers tended to give higher scores than principals, who gave higher scores than outside assessors. This may be partly a function of the preparation program received by principals and master teachers participating in this initial study. Perhaps this "halo" or tendency to give higher scores did not occur with outside assessors because of enhanced experience and understanding of STAR content and assessment process, having served as trainers in the STAR and assessment processes during the six months prior to this study. A complete report describing the study, results and conclusions/implications may be found in Teddlie, Ellett & Naik (1990).

1989-1990: A second GENOVA study was conducted during the late spring of 1990. A somewhat larger sample was used and all assessors had successfully completed all requirements of the STAR assessor certification program. Methodology, data collection and analyses procedures used in this study were the same as those employed in the initial GENOVA study. Table 18 provides a summary of the generalizability coefficients obtained in these analyses for each teaching and learning component in Performance Dimensions II, III and IV. As with the initial G-study, components in Performance Dimension I were not assessed and analyzed, since Comprehensive Unit Plans (CUPs) were not prepared and assessed. The average generalizability coefficient with the effect of all three assessors considered was .51. Results of initial generalizability analyses showed coefficients in the range of .23 to .62 for a two-person team (principal and outside assessor) and from .29 to .70 for a three-person team (adding the master teacher). These results appear somewhat lower than those obtained in the initial study, but are similar and support consistency and common perspectives across assessor types as they view classroom teaching and

learning over multiple teachers and multiple lessons. One explanation for the lower G-coefficients lies in the overall improved scores obtained by teachers. That is, for some components there was less variability in the data obtained. Teachers assessed in this study had varying levels of orientation and staff development focused on content and processes related to the STAR. Also, closer proximity to statewide implementation targets may have served to enhance teachers' performance levels. A complete report describing the study, results and conclusions/implications may be found in Teddlie, Ellett & Naik (1991). Descriptive statistics, generalizability coefficients for both indicators and components comparing two-person and three-person assessment teams and variance estimate components are included in this report, as they are in the initial G-study report, as well.

Standards-Setting Studies

June, 1990: An initial standards-setting workshop with Louisiana educators to recommend initial performance expectations for the STAR was held in June, 1990. The purpose of this workshop was to provide a highly informed ("expert") group of Louisiana educators with the results of STAR pilot research studies (1988-1989) to be used as critical information for making initial STAR performance standards recommendations for the LTIP and LTEP. In addition, the workshop served as a forum for the presentation and discussion of critical professional and program policy and implementation issues that pertained to standards recommendations.

Consistent with the recommendations of Hambleton (1978) and Shepherd (1980) on the use of several types of judges, a panel of 47 educators from various regions of Louisiana was nominated by LTIP/LTEP Coordinators giving consideration to two essential concerns: 1) knowledge and expertise in the STAR and the LTIP and LTEP; and 2) reasonable balance among panel members relative to position of employment, ethnicity, gender and other key factors. In selecting panel members, an attempt was made to assure that the majority of panelists were regular classroom teachers. All panelists nominated/selected had extensive preparation as STAR assessors and many had served during the 1989-1990 extended pilot as STAR program assistants in the assessor certification program. The LTIP/LTEP Project Director and three LSU project

coordinators organized and served as leaders for the standards-setting workshop. The outside consultant for the workshop design was Dr. Richard Jaeger, College of Education, University of North Carolina at Greensboro.

The standards-setting process, adapted from the work of Jaeger (1990), was an "iterative" one that occurred over three and one-half days of intensive workshop activity.

A variety of data were available as panelists made their recommendations from one iteration of judgments to the next. Three recommendations for a performance standard for professional, renewable certification were made for each STAR Teaching and Learning Component: 1) an initial recommendation after studying pertinent research findings and assessment indicators comprising a particular component; 2) a second recommendation after considerable discussion of the first recommendation with other panelists in small groups; and 3) a final recommendation after the results of the second recommendation with the entire group of panelists. Recommendations for "benchmark" standards for each Teaching and Learning Component were made as temporary expectations to be piloted during the first year of implementation. This panel also strongly recommended periodic and careful review and analysis of data collected under real, "high stakes" conditions in terms of these benchmark standards before making a final decisions. Ellett, Lofton, Loup, Chauvin & Evans (1990) provides a complete description of the standards-setting workshop and tasks design.

November, 1990: A follow-up activity to the initial standards-setting study was conducted in November, 1990 with a panel of all classroom teachers. Ten of these teachers were members of the "expert" panel group that set initial STAR benchmarks in June, 1990 and were selected by project staff for participation in this follow-up study. The remaining panelists, teachers who had completed a fall 1990 STAR assessment as part of LTIP or LTEP, were selected by the Department of Education. In selecting panelists, consideration was given to achieving an appropriate proportional balance by ethnicity, gender and school level. A total of 28 teachers participated in this one and one- half days of standards-setting activity. Two teachers were unable to complete all activities due to unavoidable events necessitating their early departure. In addition

to the teacher panelists, representatives of LSU LTIP/LTEP Projects and the Department of Education were in attendance, but did not participate in decision making activities. The Department of Education also provided a member of their external consultants' committee to attend as an "outside" observer.

While the purpose of this standards-setting task was not to "revisit" benchmark standards set by the initial standards-setting committee, an important aspect of this group was to review and make recommendations regarding decision making models for LTIP, LTEP (satisfactory and superior ratings). Ellett, (1990) provides a description and summary of the final set of recommendations of this panel.

External Committees: Two external committees have been convened to review, analyze, discuss and propose recommendations relative to standards-setting concerns, decision making models and elements of program implementation. The first of these two external committees was established by the Department of Education at the direction of the Louisiana Board of Elementary and Secondary Education. Six "expert" members were selected representing the LSU LTIP/LTEP Project Director, one other member from within Louisiana and four other members, one each from Texas, Michigan, Minnesota and Tennessee. This committee met in two two-day meetings during the month of November to review standards-setting issues and STAR research data.

A second external "expert" panel was convened by the LTIP/LTEP Project Director after discussion with the Department of Education. This group of consultants served as a check on the perspectives of all stakeholders and a second level review of STAR research and development and assessment results. The committee convened for an evening and a full day meeting with LSU LTIP/LTEP Project staff in Greensboro, North Carolina in late November. Department of Education representation was requested, but scheduling conflicts prevented their attendance at this meeting. Results and recommendations of these two external "expert" consultants committees are included in Ellett (1990).

Other Research Studies

Other types of investigative efforts have been included in the development and validation of the STAR. These research activities have focused on identifying educators' perceptions of the STAR, LTIP and LTEP and various aspects of utilizing large-scale teacher assessment/evaluation as an impetus for educational reform and enhancement of teaching and learning in classroom and school contexts.

A two-year extended investigation has been conducted to tap informed educators' perspectives regarding the STAR and accompanying assessment processes developed as part of the LTIP and LTEP. Also, because attention to uninformed educators' perspectives is important, as well as understanding social, political and logistical factors impacting implementation within everyday school life, qualitative studies have been conducted and are presently ongoing.

In the spring 1990, an intensive qualitative study, lasting approximately 6-8 weeks, was conducted in nine schools within a large urban school district. A followup study, encompassing a full school year and involving four schools located in different south Louisiana school districts, is currently underway in an effort to expand the field in terms of understanding the processes and interactions associated with implementation of such a large-scale teacher assessment/evaluation effort that is also focused on classroom and school improvement.

Perceptions of STAR, the LTIP and LTEP

Given that Louisiana is venturing into "new territory" by replacing lifetime teaching certificates with professional renewable certificates resulting from evaluation through an on-the-job performance assessment process, an understanding of the initial perceptions and opinions of Louisiana educators seems critical of these program are to be well-received and successful. Thus, a first effort was made in an attempt to better understand individuals' perceptions of the STAR and these programs, once they had been adequately informed. Also, information collected in this effort was used to provide formative and summative evaluation data to guide revisions, modifications, deletions and additions in the STAR and these programs during the pilot and development years, prior to statewide implementation.

During the spring of 1989 and spring of 1990, survey data were obtained from Louisiana educators who had successfully complete the STAR assessor certification program (i.e., principals, assistant principals, master teachers, instructional supervisors, college faculty and other key professional educators). These educators were asked to respond to a variety of issues related to legislation, policy, procedure and program implementation.

1988-1989: During the spring 1989, selected educators completed an eight-day pilot program designed to prepare and certify educators as STAR assessors. Principals, master teachers, instructional supervisors, college faculty and other education professionals were included in the total sample of 289. A total of 198 useable questionnaires were received, yielding a response rate of 69%.

Data collected through the questionnaires were compiled and descriptive statistics were calculated. Percentages were calculated for each survey item by scale category. In addition to percentages for the total respondent group, percentages were calculated for subgroups (e.g., principals/assistant principals, teachers, supervisors, college faculty/others). A qualitative analysis of comments provided by respondents was also conducted to identify common themes and concerns regarding the STAR and the LTIP and LTEP.

Overall, responses to the initial survey indicated that educators, regardless of position, endorse the notion of assessing on-the-job teaching performance for providing support and professional development to the beginning teacher and experienced teacher, as well as a means of granting and renewing statewide teaching certification (LTIP: 93.7% agree/strong agree; LTEP; 86.9% agree/strongly agree). Respondents also supported the team approach to observations, conferences and professional development activities. Similar support was noted for the development of a similar system for principals.

In general, comments offered by respondents strongly supported implementation of the LTIP and LTEP and the use of the STAR. However, correspondingly, they voiced concerns regarding maintenance of standards and quality in preparation programs for assessors and assessment processes for teachers, as these programs are delivered and implemented by the

Department of Education. Concerns appeared to be mainly related to the potential for shortened and rushed timelines and less-than-adequate funding provided at the local level to facilitate implementation. According to these respondents, there appeared to be a "mind set" that good things have gone awry in the past because of lack of conscientious backing by all levels involved -- state and local educators and policy makers. A complete description of this study, results and conclusions/implications may be found in Chauvin & Ellett (1990a).

1989-1990: A second research effort was conducted in the spring of 1990 with a larger sample to continue to assess perceptions held by informed Louisiana educators regarding the STAR and the LTIP and LTEP. A questionnaire was developed to represent a revised version of the instrument used in the prior, preliminary study described above. This instrument was improved and expanded to more accurately reflect concerns and issues which had been identified as a result of continuing research and developmental activities since August, 1988. Individual items were revised and additional items were written to reflect information obtained as a result of: 1) ongoing analysis of each legislative document; 2) observations and results of pilot activities; and 3) questions, issues and concerns raised by educators statewide.

A random sample of 1200 educators drawn from some 2500 participants who had successfully completed a seven-day Professional Development Program to Certify STAR Assessors during the 1989-1990 extended pilot year. Participants were selected from all 66 school districts in Louisiana and included master teachers, principals and assistant principals, instructional supervisors, college faculty and other education professionals such as Department of Education personnel. Of the 1200 questionnaires mailed, 920 useable instruments were obtained, resulting in a return rate of 76%. Methodology, data collection and analyses procedures were the same as those employed in the initial study conducted during the spring of 1989.

Very similar results were obtained in this followup study to those obtained with the smaller sample in the initial investigation. Overall, it appears that educators, regardless of current position, continue to endorse on-the-job performance assessment for both beginning and experienced teachers (Total respondents - LTIP: 89.5% agree/strongly agree; LTEP: 67.2% agree/strongly agree).

However, support for LTEP does not seem to be as strong as that evidenced for LTIP, or as strongly as communicated in the 1989 survey. While there appeared to be some disparity of responses between respondent groups, there did not appear to be many substantial differences.

Other similar findings to that first evidenced in the initial survey include support for the team approach to assessment, conferences and professional development activities. While individual strongly supported the process as a professional obligation (>91% agree/strongly agree), educators appeared sensitive to additional out-of-school time requirements without additional compensation. With respect to the development of a "Comprehensive Unit Plan", which represents an assessment component targeting reflective planning practices, strong support for inclusion of this requirement was evidenced for both LTIP (89.7% agree/strongly agree) and LTEP (75.1% agree/strongly agree). Interestingly, this requirement was retained for LTIP implementation, but discarded for LTEP statewide implementation. Results revealed strong support for the STAR as an assessment system that is fair and impartial (74.1% agree/strongly agree) and one that is useful in developing professional improvement plans (90.9% agree/strongly agree). Further, results indicated strong agreement supporting the development of corresponding staff development programs (96.5% agree/strongly agree). A complete description of this study, results of descriptive analyses (quantitative and qualitative) for the total group and indentified subgroups, and conclusions may be found in Chauvin & Ellett (1990b).

Results from both survey efforts seem to support several conclusions. While informed educators, statewide, seem to strongly support the STAR assessment process for both the LTIP and LTEP, survey results revealed concerns pertaining to policy decisions, confidentiality of assessment results, and due process provisions. Responses to open-ended questions revealed concerns relative to maintenance of standards and quality in the preparation programs for STAR assessors and assessment process for teachers, as the programs are delivered and implemented by the Department of Education. In particular, concerns seemed to focus on "time" and "money" issues. Also, numerous responses revealed concerns over bureaucratic and political interferences that have the potential to block successful implementation and educational improvement in Louisiana.

Qualitative Studies

Nine School Study: This study was conducted as a preliminary investigation of the links between school context variables and the resultant receptivity of school faculty/staff to the STAR and the LTIP and LTEP. The study utilized data collected over a period of three months in the spring of 1990 as part of a criterion-related validity study of the STAR. This trial study of the total STAR assessment process was completed in nine schools in one urban Louisiana school district. Schools included in this study varied on two dimensions: 1) grade level of students served: elementary, middle, high; and 2) socioeconomic status (SES) of student body: low, middle, high. Two SES variables were used in the selection process for the nine schools: percentage of mothers who had some college education and percentage of fathers with white collar jobs. Once schools were categorized as low, middle or high SES, one school from each category was randomly selected to participate in the study. A total of 54 teachers (6 teachers in each school: two student teachers, two beginning/first year teachers, and two experienced teachers) participated in a three month (March through May) "abbreviated" version of the STAR model assessment year.

Data for this study included STAR assessment results on participating teachers, interviews conducted with STAR team members, logbooks completed by participants, observation notes taken during STAR post-assessment conferences, participant survey responses and assertions regarding their beliefs about effective teaching, LTIP/LTEP and the STAR. Interview and observation data were collected by the nine members of the university research team also serving as "outside assessors" for the STAR assessment teams. Educator assertions and some interview data were collected by teacher researchers in the schools. In addition, a three-person experienced qualitative research team from an out-of-state university collected data over two one-week periods during the latter part of the research period and conducted qualitative analyses from a purely external or "outsider" perspective.

Throughout the three months of the study the nine teacher researchers (one in each school) wrote down "assertions" reflecting casual remarks, comments, and specific statements obtained

from conversations with other educators employed at the school. These data were reported by the teacher researchers in the form of initial assertions. These initial assertions represented fairly specific statements, with little inferencing on the part of the teacher researchers. These initial assertions were directly tied to educators' comments, either through direct quotes or by paraphrased accounts. Initial assertions were then grouped into categories, reflective of emergent themes in the data. General statements representing these categories then formed the basis for higher "level two" assertions. Finally, these higher assertions resulted in the emergence of a few theory-based assertions from the entire data set. Complete reports of various qualitative investigations may be found in Claudet, Chauvin & Loup (1991), LeMaster, Tobin & Bowen (1990) and Tobin (1990b). Appendix B summarizes highlights of results and conclusions from this qualitative investigation (Chauvin, 1991).

STAR Professional Development Project (Four Schools): Currently a school year-long study is being conducted to learn more about the influences and impact of the STAR as a staff development and professional development framework on the "everyday life" in classrooms and schools. Each school, representing different school districts have been included based on agreement by school personnel to commit to this intense and long-term investigative effort. Selection of the schools was made based on a number of factors, including demographics, student population, grade level of students served. Two schools are rural primary elementary schools serving a mixed ethnic population, a third school serves students enrolled in middle grades with a population of 60% minority. The fourth school included in the current study is a secondary school (grades 8-12) serving 100% minority in a rural setting. Three of the schools have experienced past concerns regarding student attendance and below level student achievement.

Teachers are maintaining journals in which they record "critical incidents", thoughts, observations and concerns regarding the use of the STAR as they are involved in related staff development and professional development activities. Building administrators are also maintaining journals. Classroom-based assessment data for each participating teacher is being collected using the STAR. In addition, long-term student achievement data is being collected pre- and posttest

using standardized measures, as well as student surveys of their perceptions of the learning environment are also being collected pre- and posttest.

University project staff, serving as external change agents, are maintaining journals and narrative accounts of observations, "critical incidents" and interview data during the year-long study. Results of these efforts will be forthcoming upon the conclusion of the 1990-1991 pilot year.

Alternative Applications

As already mentioned, the STAR was developed as a professional, contextually-based assessment and decision making framework and represents much more than a teacher evaluation "checklist". Thus, with the STAR, assessors must attend to the interactive nature of teaching and learning as it actually occurs, since effective teaching is conceptualized in the STAR as a professional activity and "adaptive dance" that targets the enhancement of student learning within a complex social learning environment (Ellett, 1990a).

To date, the STAR has been used with few exceptions to applicability to context in approximately 7000 classroom assessments. Two years of development, piloting and validation have been completed with the STAR. Considered, collectively, these studies support the psychometric properties of the STAR (validity and reliability)....but more importantly, they suggest that the STAR may be used as comprehensive, classroom-based system in other learning contexts (e.g., higher education classrooms), knowledge assessment of teachers' knowledge of content, pedagogy and curriculum (e.g., portfolio and semi-structured interview assessment of preservice and inservice teachers) and as measurement of other perspectives of classroom contexts (e.g., a comprehensive assessment system of "learning" environment characteristics).

Higher Education Classroom Contexts: During the summer of 1990 an initial pilot of the STAR was conducted to explore its applicability as an alternative to the traditional model of using student evaluations of instruction in higher education contexts. This initial pilot was conducted with experienced assessors in classes taught by graduate teaching assistants (GTAs) in six different contexts: mathematics, chemistry lab, biology lab, speech, English and psychology. Results of the

initial pilot indicated that the STAR was adaptable to higher education contexts (Evans & Ellett, 1990). In the fall of 1990, an expanded pilot was conducted with a larger number of GTAs in a wider variety of college classes covering twelve different content areas and 25 classrooms at a large research university. Both quantitative and qualitative data were collected and analyzed. Results of this expanded pilot showed considerable variability between classes in the components related to effective teaching and learning. Results of classroom observations in the component of "Thinking Skills" indicate that this is a critical need area in terms of enhancing learning for students. Comparisons between 25 higher education contexts and data collected on 6000 elementary and secondary classrooms were made and are provided in Table 19. GTAs performed at lower levels for each Teaching and Learning Component, except for "Classroom Routines". Of note, was the significantly lower percentage of evidence for teaching thinking skills in the higher education classrooms (12% "acceptable") than elementary and secondary classrooms (22% "acceptable"). A complete description of activities completed to date in this extended pilot, results, conclusions and recommendations may be found in Evans & Ellett (1991).

Continued efforts in the application and adaptation of the STAR to higher education context are currently focused on completing additional assessment in an even wider range of classroom contexts (e.g., large group lecture hall settings, complex laboratory contexts, etc.). Work is being started on the development of a draft version of the STAR adapted for higher education contexts. In keeping with the focus on support and professional development, research and development activities being conducted in spring 1991 are focusing on the use of the STAR in assessment, post-assessment reflective practice conferences and professional development activities with a select group GTAs volunteers. Results of these efforts are forthcoming in a series of technical reports to be completed at the conclusion of the 1990-1991 pilot year.

Knowledge Assessment and Reflective Practice: An initial probe into the assessment of teachers' knowledge of content, pedagogy and curriculum using comprehensive planning and reflective practice through semi-structured interview assessment was conducted with a small select group of preservice teachers during the extended pilot year. Case studies of student teachers and

assessment of their knowledge of context specific content structure and pedagogy has been conducted and is reported in Hill, Lee & Lofton (1990). Results point to preservice teachers' ability to plan content and learning activities relative to form, but not substance. That is, preservice teachers participating in this initial investigation could adequately plan a body of content "for a professor", but had much difficulty in planning adequately for a specific group of students in a way that reasonably accommodated developmental and ability levels and individual learning needs. In addition, to adequately structure content to meet individual learning needs of students, preservice teachers evidenced difficulty with planning appropriate breadth and depth of content, consideration of curriculum scope and sequence, as well as consideration of learning outcomes versus "things to do".

Followup investigations are currently being designed and implemented using Comprehensive Unit Plans (CUPs), semi-structured interviews and subsequent classroom observations with first-year, beginning teachers and experienced teachers to explore broader assessment of teaching and learning in terms of knowledge of content, pedagogy and curriculum and important abilities related to professional reflective practice.

Alternative Assessment of Learning Environments: Though the STAR was originally designed to meet the legislative mandates of the Louisiana Teaching Internship Law (1984) and the Children First Act (1988), it represents far more than yet another teacher evaluation system. It was designed as an integrated, comprehensive assessment of the total learning environment. In this sense, it seems to offer an alternative to more traditionally used and more narrowly focused measures of students' perceptions of the psychosocial elements of classrooms that have characterized the past two to three decades of research on classroom environments. The STAR is focused on not only teaching effectiveness, but the nature of social interactions in the classroom and student "learning" as well. This focus provides support for its utility as a comprehensive, "in situ" measure of elements of the total "learning" environment...not the more narrow psychosocial properties of classrooms generally obtained on paper-and-pencil student perceptions measures. Early studies of differences among classrooms, effectiveness of teaching and students' learning

suggest the STAR can be a viable addition to the measurement of learning environments that can move our understandings through future research beyond the past two decades of students' perceptions of the classroom context. Thus, an alternative application of the STAR as an appropriate measurement in future studies of learning environments is being initially explored. Preliminary findings, conclusions and implications for further study are forthcoming in Loup, Ellett and Chauvin (1991).

Conclusion

Though the STAR was originally designed to meet the legislative mandates of the Louisiana Teaching Internship Law (1984) and the Children First Act (1988), it represents far more than yet another teacher evaluation system. It was designed as an integrated, comprehensive system of teaching and learning that encompasses the holistic nature of context and interactions occurring within any lesson unlike other large-scale teacher evaluation systems. Thus, the STAR is clearly a part of a new generation of assessment systems.

Research findings offer convincing evidence that the STAR is a system that can validly and reliably assess not only effective teaching, but also make inferences about student learning in a wide variety of classroom contexts. Research findings also support the ability of the STAR to differentiate "superior" from "typical" teachers, and assess newer and important areas such as, teaching students higher-order thinking skills and structuring content and pedagogical knowledge. Thus, the STAR seems to offer much promise of contributing to a field of performance-based teacher assessment as part of a "new generation" of assessment systems that "puts the light on the learner".

Although two years of extensive research and development efforts have been exerted prior to statewide implementation of the LTIP and LTEP using the STAR, continued investigations are necessary to explore the measurement characteristics of this system as they are evidenced under real "high stakes" conditions. For example, preliminary data based on fall 1990 assessments show that presently principals are "inflating" assessment decisions at approximately 2 1/2 times the rate as master teachers and outside assessors also serving on assessment teams. Also, it should be

noted that under current implementation procedures, master teachers have been assigned full time to assessment teams and are not currently teaching in a classroom. Thus, the present model, which differs from the intent and language of legislation, as well as the assessment model designed and piloted during 1988-1989 and 1989-1990, and in essence, includes two "outside" assessors. The principal is currently the only "in-building" assessor for the LTIP and LTEP. This and other features of actual implementation practices that are different from piloted processes are being observed and analyzed in terms of impact on the measurement integrity and potential for effecting positive professional growth and improvement.

Another area of concern in need of continued attention will be the effect of "assessor drift" over time and the influence of update sessions and "recertification" requirements for STAR assessors. Presently, the Department of Education has not finalized plans to address these areas. Recommendations submitted by the developers are currently being considered. In any case, future investigations are warranted.

Because the STAR and legislation underwriting the LTIP and LTEP places strong emphasis on formative and summative use of assessment data for the purpose of professional growth and development, it will be important to continue investigations and on-going study of the effects of the STAR, assessment processes (LTIP and LTEP) and corresponding professional development activities on positive change in teachers' professional practice, students' learning and classroom/school learning environments. One such effort currently underway, as mentioned in an earlier section of this paper, is an intensive study (quantitative and qualitative) involving four schools.

Utilizing data from assessments conducted under real conditions involving approximately 8000 experienced teachers and 1500 beginning teachers, a series of reliability and validity studies will be conducted as part of the 1990-1991 fiscal year. Analyses of these data will offer important information that will serve to guide future developments of the STAR and related assessment issues (e.g., standards-setting).

Finally, while the STAR has been developed and validated for use with classroom teachers, the Children First Act (1988) also mandates on-the-job performance assessment of special category teachers (i.e., school librarians, guidance counselors, speech-language pathologists and assessment teachers) in a manner consistent with processes established for regular classroom teachers. Developmental work has begun this spring (1991) to design adaptations of the STAR and development of assessment processes appropriate for each of these special categories. Thus, similar developmental and validation studies, as well as statewide pilot activities, will be necessary relative to these adaptations of the STAR.

In conclusion, during the 1988-1989 and 1989-1990 pilot years and continuing into the 1990-1991 year of initial implementation, the development of the STAR for the LTIP and LTEP has enthusiastically focused the "light on the learner" as part of a "new generation" of teacher assessment systems. As we continue to "focus the light", additional concerted efforts from a variety of perspectives (e.g., research and development, state-level implementation, local district support, and individual professional commitment) will be continued to be needed so that students and their learning may be enhanced in Louisiana's classrooms.

References

Berliner, D. (1986, September/August). In pursuit of the expert pedagogue. Educational Researcher. 5-13.

Brophy, J. (1986, October). Teacher influences on student achievement. American Psychologist. 1069-1077.

Capie, W. & Ellett, C. D. (1982, March). The effects of assessment demand characteristics on the dependability of teacher performance measures. Paper presented at the annual meeting of the American Educational Research Association, New York.

Capie, W., Ellett, C. D. & Cronin, L. (1985). Assessing meritorious teacher performance: Reliability, decision-making and standards setting procedures. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.

Capie, W., Tobin, K., Ellett, C. D. & Johnson, C. (1981, April). The dependability of job performance rating scales for making classification decisions. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Chauvin, S. W. (1991, January). A qualitative study of implementation of the STAR in nine schools. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, Texas.

Chauvin, S. W. & Ellett, C. D. (1990a). Initial perceptions of educators in Louisiana regarding the Louisiana Teaching Internship and Statewide Teacher Evaluation Programs. Technical Report Number 3, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Chauvin, S. W. & Ellett, C. D. (1990b). A followup survey of Louisiana educators' perspectives about the Louisiana Teaching Internship and Statewide Teacher Evaluation Programs. Technical Report Number 9, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Chauvin, S. W., Ellett, C. D., Loup, K. S. & Slan, D. F. (1990, April). Inserting student learning in the teacher assessment paradigm: Results of an initial pilot test of the Louisiana STAR. Paper presented at the annual meeting of the American Educational Research Association, Boston, Massachusetts.

Claudet, J. G. & Ellett, C. D. (1990, March). Review of pertinent literature and theoretical foundations of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 8, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Claudet, J. G. & Ellett, C. D. (1990). Updated review of research findings and theoretical foundations of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 19, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Claudet, J., Hill, M. F., Ellett, C. D. & Naik, N. (1990). Summary of statewide assessment from the 1989-1990 extended pilot of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 11, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Crick, J. E., & Brennan, R. L. (1983). GENOVA: A general purpose analyses of variance system. Iowa City: American College Testing Program.

Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart & Winston.

Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnan, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.

Ellett, C. D. (1990a). A new generation of classroom-based assessments of teaching and learning: Concepts, issues and controversies from pilots of the Louisiana STAR. College of Education, Louisiana State University.

Ellett, C. D. (1990b). Brief summary of results of fall 1990 standards-setting workshop with Louisiana teacher panelists. Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Chauvin, S. W., Loup, K. S. & Naik, N. S. (1990a). Summary of statewide assessment data from the 1989 pilot of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 2, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Chauvin, S. W., Loup, K. S. & Naik, N. S. (1990b). Content verification of assessment indicators by selected Louisiana educators' certified in the use of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 10, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Garland, J. & Logan, C. (1987). Content classification, synthesis and verification of eight large-scale teacher performance instruments. Research report, Teaching Internship Project, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Lofton, G., Loup, K. S. & Chauvin, S. W. (1990). Results of an initial iterative standards-setting task for certification of teachers using the System for Teaching and learning Assessment and Review (STAR): 1990 Revised Edition. Technical Report Number 18, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Loup, K. S. & Chauvin, S. W. (1990). System for Teaching and learning Assessment and Review (STAR): 1990 Revised Edition. Assessment Manual and Interpretive Guide, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Loup, K. S., Chauvin, S. W., Lofton, G. & Naik, N. (1990). An initial investigation of the validity of the System for Teaching and learning Assessment and Review for differentiating "superior" teachers from other teachers. Technical Report Number 14, Louisiana Teaching and Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Loup, K. S., Chauvin, S. W. & Naik, N. (1990). Initial exploratory factor analyses of assessment indicators comprising the System for Teaching and learning Assessment and Review (STAR). Technical Report NUmber 5, Louisiana Teaching and Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Loup, K. S., Chauvin, S. W. & Naik, N. (1990). An initial investigation of the criterion-related validity of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 6, Louisiana Teaching and Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Loup, K. S., Chauvin, S. W. & Naik, N. (1990). A study of the generalizability and dependability of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 16, Louisiana Teaching and Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Ellett, C. D., Lofton, G., Loup, K. S., Chauvin, S. W. & Evans, R. L. (1990). Results of an initial iterative standards-setting task for certification of teachers using the System for Teaching and learning Assessment and Review (STAR). Research report, Louisiana Teaching and Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Evans, R. L. & Ellett, C. D. (1990, July). An initial application of the System for Teaching and learning Assessment and Review (STAR) to the assessment of teaching and learning in selected undergraduate classrooms at Louisiana State University. Technical Report Number 1, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Evans, R. L. & Ellett, C. D. (1991, January). Application of a direct observation system to assess teaching and learning in higher education contexts. Paper presented at the annual meeting of the Southwest Education Research Association, San Antonio, Texas.

Hambleton, R. K. (1987). On the use of cut-off scores with criterion-referenced tests in instructional settings. Journal of Educational Measurement, 15(4), 277-290.

Hill, F., Lee, A., & Lofton, G. (1991). Assessing the relationship between reflective practice, content knowledge and teaching effectiveness of student teachers. Paper presented at the annual meeting of the American Edcuational Research Association, Chicago, Illinois.

Jaeger, R. M. (1990). Setting standards on teacher certification tests. In J. Millman and L.D. Hammond (Eds.) The new handbook of teacher evaluation: Assessing elementary and secondary school teachers. Newbury Park, California: Sage Publications.

LaMaster, S., Tobin, B. & Bowen, C. (1990, June). Making sense of an innovation: A qualitative sutdy of the pilot implementation of LTIP/LTEP in three schools. College of Education, Florida State University, Tallahassee, Florida.

Lofton, G., Ellett, C. D., Chauvin, S. W., Loup, K. S. & Claudet, J. G. (1990). Continuing investigations of the criterion-related validity of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 13, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Loup, K. S., Chauvin, S. W., Ellett, C. D. & Naik, N. (1990). Confirmatory factor analyses of assessment indicators comprising the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 12, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Loup, K. S., Chauvin, S. W., & Ellett, C. D. (1990). Results of a survey to corroborate qualitative findings from pilot implementation of the System for Teaching and learning Assessment and Review (STAR) in nine schools. Technical Report Number 17, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Medley, D. L. & Mitzel. (1963). Measuring classroom behavior by systematic observation. In N.C. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally.

Porter, A. C. & Brophy, J. (1988). Synthesis of research on good teaching: INsights form the work of the institute for research on teaching. Educational Leadership, 74-85.

Performance Assessment Systems, Inc. (1984). A study of the genralizability of the Teacher Assessment and Development System (TADS) MTP form. Athens, GA: Author.

Shepard, L. (1980). Standard setting issues and methods, Applied Psychological Measurement, 4(4), 447-467.

Shulman, L. S. (1986, February). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. Harvard Educational Review, 57(1), 1-22.

Teddlie, C., Ellett, C. D., & Naik, N. (1990). A study of the generalizability and dependability of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 4, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Teddlie, C., Ellett, C. D., & Naik, N. (1991). A study of the generalizability and dependability of the System for Teaching and learning Assessment and Review (STAR). Research report, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Tobin, K. (1990a). A summary report of "experts" external reviews of the System for Teaching and learning Assessment and Review (STAR). Technical Report Number 18, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

Tobin, K. (1990b). A qualitative study and synthesis of issues and concerns in the Louisiana Teaching Internship and Teacher Evaluation Programs: Learning from pilot program implementation in nine schools. Technical Report Number 15, Louisiana Teaching Internship and Teacher Evaluation Projects, College of Education, Louisiana State University, Baton Rouge, Louisiana.

APPENDIX A-1

STAR

System for Teaching and Learning Assessment and Review

PERFORMANCE DIMENSION I: PREPARATION, PLANNING AND EVALUATION (32)a

TEACHING AND LEARNING COMPONENTS

Component #c		
1	A.	Goals and Objectives (6)b
2	B.	Teaching Methods and Learning Tasks (6)
3	C.	Allocated Time and Content Coverage (4)
4	D.	Aids and Materials (5)
5	E.	Homework (4)
6	F.	Formal Assessment and Evaluation (7)

PERFORMANCE DIMENSION II: CLASSROOM AND BEHAVIOR MANAGEMENT (28)

TEACHING AND LEARNING COMPONENTS

7	A.	Time (8)
8	B.	Classroom Routines (4)
9	C.	Student Engagement (1)
10	D.	Managing and Task-Related Behavior (6)
11	E.	Monitoring and Maintaining Student Behavior (9)

PERFORMANCE DIMENSION III: LEARNING ENVIRONMENT (16)

TEACHING AND LEARNING COMPONENTS

12	A.	Psychosocial Learning Environment (12)
13	B.	Physical Learning Environment (4)

PERFORMANCE DIMENSION IV: ENHANCEMENT OF LEARNING (64)

TEACHING AND LEARNING COMPONENTS

14	A.	Lesson and Activities Initiation (10)
15	B.	Teaching Methods (6)
16	C.	Aids and Materials (8)
17	D.	Content Accuracy and Emphasis (7)
18	E.	Thinking Skills (11)
19	F.	Clarification (5)
20	G.	Pace (3)
21	H.	Monitoring Learning Tasks and Informal Assessment (6)
22	I.	Feedback (4)
23	J.	Oral and Written Communication (4)

---

a Number of Assessment Indicators Comprising Performance Dimension
b Number of Assessment Indicators Comprising Teaching and Learning Component
c "Component #" identifies components referred to in the tables in Appendix A.

TEACHING AND LEARNING COMPONENT II.A: TIME

ASSESSMENT INDICATORS	ANNOTATION	NOTES/CLARIFICATION
II.A.1 Learning activities begin promptly	This indicator focuses on the *beginning of the lesson.* Learning activities should begin with little time spent on organizational activities such as roll taking and distributing materials and supplies. The efficiency with which organizational activities are handled is always a concern. **IF A SIGNIFICANT AMOUNT OF TIME IS WASTED AT THE BEGINNING OF THE LESSON, THE INITIAL USE OF TIME IS UNACCEPTABLE.**	
II.A.2 Expectations for maintaining and completing timelines for tasks are communicated to students.	As initial tasks begin and as tasks change throughout the lesson, the teacher should clearly communicate to students when tasks are to be completed. Cautions about wasting time and informing students about the persistence needed to complete tasks on time are elements of effective communication of expectations. **IF THE TEACHER DOES NOT ADEQUATELY COMMUNICATE THESE EXPECTATIONS TO STUDENTS, THE USE OF TIME AVAILABLE FOR LEARNING IS UNACCEPTABLE.**	

Appendix B

SYSTEM FOR TEACHING AND LEARNING ASSESSMENT AND REVIEW STAR

Louisiana Teaching Internship and Statewide Teacher Evaluation Program (LTIP/LTEP)

NINE SCHOOL STUDY

Spring, 1990

I. EDUCATORS' BELIEFS ABOUT TEACHING

- The degree to which personal beliefs about teaching and learning are congruent with key elements of the STAR influence one's acceptance of the STAR as a valid system.
  - Content coverage versus students' learning
  - Activity versus learning
  - Emphasis on excuses versus opportunities
  - Attitude toward professional development

II. EDUCATORS' BELIEFS ABOUT LTIP/LTEP AND THE STAR

- [Despite careful planning and research], "its ultimate success will hinge upon the attitudes and commitment of all persons involved in its' implementation." (Participant/Observer comments)
  - Lack of information, rumors and much "misinformation" resulted in many teachers being fearful of the STAR and the LTIP/LTEP process. They were also mistrustful of pilot implementation and use of the STAR in these processes.

    However, where information was shared in a positive and professional manner, teachers appeared comfortable and positive.
  - For many, perceptions did not allow for a pilot period; implementation began with legislation.
  - Initial view of "getting rid of bad teachers" versus professional development and collaboration for all educators focused on enhancing students' learning
  - "Dog and pony show" versus enhanced professional practice (power in the "getting ready")
  - Everyday practice versus a certification/licensure procedure
  - Confusion between employee issues of tenure and employment and state

certification/licensure)

* Opposition to violation of "lifetime certificate" (sacred norm) and not content of STAR (viewed as useful in professional development)

* Focus on hindrance factors associated with implementation (e.g. time, money, scheduling and other "extra effort" concerns)

III. PREPARATION AND PLANNING (COMPREHENSIVE UNIT PLAN)

TEACHER'S PERSPECTIVE:

* Planning appears to thought of in terms of "things to do" to fill the time available, rather than as "steps" that lead to accomplishmnet of "what students are to learn and know".

* Teachers seem to have much difficulty in structuring content. While little difficulty was observed in discussing rationales for activities, discussions of rationales for content order and structure was either difficult for teachers OR content was not clearly included.

* Teachers seem to have much difficulty in planning for student needs and abilities (accomodating individual differences).

* Teachers did not understand how to use conten[illegible] STAR Performance Dimension I: Preparation, Planning and Evaluation, to structure a Comprehensive Unit Plan. Teachers expressed a desire for samples, formats and models from which they could copy. They expressed much difficulty in coping with open-ended possibilities of structuring a comprehensive plan for a given body of content and a particular group of students.

* Despite difficulties experienced in structuring a Comprehensive Unit Plan (CUP), teachers who did complete such a plan appeared, and self-reportedly, were more prepared and organized than when a CUP was not constructed.

ASSESSOR'S PERSPECTIVE:

* The Comprehensive Unit Plan helps to clearly establish the teaching and learning context to be observed. Assessors more clearly knew what to expect, than with a brief daily lesson plan.

* Teachers appeared to be more comfortable in lessons resulting from the preparation of a CUP, and activities during the lesson appeared to be more organized, efficient and effective in terms of student involvement than when daily lesson plans were used.

* Preparation of a Comprehensive Unit Plan appears to enhance subsequent success in the teaching and learning process during lessons.

IV. IMPLEMENTATION WITHIN EVERYDAY SCHOOL LIFE

* Influenced by the attitudes and levels of commitment of the principal and master teacher

- Sets the tone and contributes to investment of commitment by teachers
- Initial bearers of information and/or misinformation
- Where there was positive support and commitment, increased evidence of teachers including new ideas and striving for improvement was observed.

* Introduction of process was met with anxiety and apprehension, which subsided with time and positive/successful experiences

* Positive results in terms of scheduling and professional outcomes hinges heavily on commitment to clear and comprehensive planning

* Students noticed differences in lessons that were observed and those typical of everyday ("Class is better when you are here.")

V. ROLE OF PROFESSIONAL DEVELOPMENT CONFERENCE

* Focus on "scores" versus professional growth and collaboration

* Physical and pyschosocial environment

* Participation of assessment team members (including assessee)
* Understanding of participants' roles in a professional development conference

  - Expectations
  - Preparation and planning
  - Participation
  - Commitment and change

TABLE 1

Percentage of Maximum Possible for Teaching/Learning Components for each Dimension of the STAR Teaching/Learning Components (118 indicators) (N = 969)

TEACHING/LEARNING COMPONENTS		# of Indicators	Maximum Possible	% of Maximum
PERFORMANCE DIMENSION II CLASSROOM BEHAVIOR AND MANAGEMENT				
A.	Time	8	7752	73.41
B.	Classroom Routines	4	3876	81.84
C.	Student Engagement	1	969	47.47
D.	Managing Task-Related Behavior	7	6783	62.14
E.	Monitoring/Maintaining Student Behavior	10	9690	67.46
PERFORMANCE DIMENSION III LEARNING ENVIRONMENT				
A.	Psychosocial	15	14535	72.73
B.	Physical	5	4845	88.69
PERFORMANCE DIMENSION IV ENHANCEMENT OF LEARNING				
A.	Lesson Activities Initiation	10	9690	50.23
B.	Teaching Methods	5	4845	71.04
C.	Sequence/Pace	5	4845	65.59
D.	Aids and Materials	10	9690	72.06
E.	Content Accuracy/Emphasis	8	7752	65.26
F.	Thinking Skills	11	10659	38.83
G.	Clarification	5	4845	67.47
H.	Monitoring Learning Tasks/ Informal Assessment	6	5814	54.09
I.	Feedback	4	3876	53.02
J.	Oral/Written Communication	4	3876	94.66

TABLE 2
Percentage of Maximum Possible for Teaching and Learning Components for Each Dimension of the STAR
(Indicators = 108)
(N = 5720)

TEACHING AND LEARNING COMPONENTS		# of Indicators	Maximum Possible	% of Maximum
PERFORMANCE DIMENSION II: CLASSROOM AND BEHAVIOR MANAGEMENT				
A.	Time	8	43,784	72.39
B.	Classroom Routines	4	21,892	74.17
C.	Student Engagement	1	5,473	36.87
D.	Managing Task-Related Behavior	6	32,838	48.48
E.	Monitoring and Maintaining Student Behavior	9	49,257	54.21
PERFORMANCE DIMENSION III: LEARNING ENVIRONMENT				
A.	Psychosocial	12	65,676	66.40
B.	Physical	4	21,892	88.03
PERFORMANCE DIMENSION IV: ENHANCEMENT OF LEARNING				
A.	Lesson and Activities Initiation	10	54,730	34.45
B.	Teaching Methods and Learning Tasks	6	32,838	58.64
C.	Aids and Materials	8	43,784	61.78
D.	Content Accuracy and Emphasis	7	38,311	49.14
E.	Thinking Skills	11	60,203	21.56
F.	Clarification	5	27,365	54.28
G.	Pace	3	16,419	58.02
H.	Monitoring Learning Tasks and Informal Assessment	6	32,838	43.15
I.	Feedback	4	21,892	33.22
J.	Oral and Written Communication	4	21,892	94.70

Table 3

Summary of Intercorrelations Between STAR Teaching and Learning Components and Subscales of the My Class Inventory, Achievement Gain Index and Class Engagement Rate. (n=30 Elementary Classrooms)

My Class Inventory Subscales	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Cohesiveness	.28	.01	.44*	.29	.25	.05	.06	-.05	.24	.11	.33	-.03	-.08	.27	.25	.00	.11
Friction	-.01	-.28	-.11	.02	-.09	-.24	-.07	-.13	.10	-.14	.04	-.20	-.22	-.18	.09	-.09	.08
Difficulty	-.30	-.33	-.66**	-.22	-.28	-.18	-.10	-.02	.09	-.20	-.10	-.05	-.02	-.31	-.05	-.21	.08
Satisfaction	.26	.52**	.42*	.18	.39	.38	.06	.26	.22	.40*	.29	.28	.23	.29	.26	.53**	.19
Competitiveness	.37	.23	.10	.21	.33	.39	.45*	.25	.49**	.32	.60**	.47*	.37	.36	.52*	.45*	.20
Achievement Gain Index	.38*	.16	-.02	.29	.23	.25	.12	.40*	.15	.15	.25	.34	.51**	.34	.35	.33	.18
Engagement Rate	.53**	.54**	.61**	.48**	.59**	.51*	.38*	.30	.53**	.56**	.56**	.21	.15	.40*	.49**	.51*	.57**

*p<.05
**p<.01

Table [illegible]

Summary of Intercorrelations Between STAR Teaching and Learning Components and Subscales of the Learning Environment Inventory, Achievement Gain Index and Class Engagement Rate. (n=36 Secondary Classrooms)

	STAR Teaching and Learning Components																
Learning Environment Inventory	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Cohesiveness	-.08	-.09	.25	.26	.02	.16	.03	.08	.08	.10	-.07	-.08	.02	.05	.11	.18	-.08
Friction	-.25	-.26	-.24	-.08	-.23	-.27	-.24	-.21	-.15	-.24	-.17	-.28	-.34*	.03	-.22	-.19	-.31
Difficulty	.37*	.39*	-.01	.08	.24	.23	.31	.15	.27	.20	.18	.30	.20	.06	.19	.25	.34*
Satisfaction	-.27	-.20	.16	-.08	-.12	-.03	-.30	-.07	-.20	-.12	-.20	-.15	.01	-.21	-.08	-.14	-.16
Competitiveness	.07	-.01	-.18	.16	.11	-.07	-.15	-.05	.07	-.06	.05	-.04	-.06	-.02	.13	-.05	-.20
Achievement Gain Index	.08	.13	.21	.09	.17	.10	-.02	.17	.04	.13	-.11	.17	.30*	.09	.10	.02	-.06
Engagement Rate	.71**	.73**	.50*	.25	.58**	.73**	.72**	.53**	.44**	.57**	.54**	.66**	.41**	.44**	.42**	.49**	.81**

*p<.05
**p<.01

Table 5

Summary of Intercorrelations Between STAR Teaching and Learning Components Achievement Gain Index and Class Engagement Rates for All Classrooms (n=66)

	STAR Teaching and Learning Components																
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Achievement Gain Index	.19	.14	.16	.30*	.24	.19	.09	.29*	.21	.19	.11	.27*	.40**	.21	.30*	.19	.04
Class Engagement Rate	.64**	.62**	.53**	.40**	.59**	.57**	.65**	.47**	.54**	.57**	.57**	.55**	.35**	.42**	.49**	.51**	.79**

Table 6

Summary of Intercorrelations Between Subscales of the My Class Inventory (MCI), the Learning Environment Inventory (LEI), Achievement Gain Index and Class Engagement Rate Indices for Elementary (n=30) and Secondary (n=36) Classrooms.

MCI and LEI Subscales	Class Engagement Rate		Achievement Gain Index	
	Elementary	Secondary	Elementary	Secondary
Cohesiveness	.26	-.01	.43*	.34
Friction	-.09	-.28	.32	-.22
Difficulty	-.40	.38	-.21	-.04
Satisfaction	.51**	-.16	.19	.23
Competitiveness	.36	-.24	.03	-.07
Achievement Gain Index	.02	-.05	---	---

* P<.05

TABLE 7
Summary of Intercorrelations Between STAR Teaching and Learning Components and Subscales of the My Class Inventory (n=24 Elementary Classrooms)

	STAR Teaching and Learning Components																
My Class Inventory Subscales	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Cohesiveness	.39	.44*	.34	.43*	.53**	.17	.26	.28	.31	.20	.25	.15	.42*	.45*	.28	.27	.34
Friction	-.26	-.17	.29	.08	-.13	.21	-.19	.30	.27	.25	.18	.06	.19	.08	-.25	.01	.11
Difficulty	.06	.23	.17	.22	.30	.28	-.08	.26	.22	.14	.07	.21	.08	.28	.03	.31	-.18
Satisfaction	.19	.14	.32	.18	.29	.00	.43*	.18	.12	.24	.36	.19	.25	.22	.08	.06	-.01
Competitiveness	-.01	.31	.39	.20	.27	.06	-.10	.29	.34	.25	.36	.20	.35	.13	.17	.23	-.16

TABLE 8

Summary of Intercorrelations Between STAR Teaching and Learning Components, Achievement Gain Index and Class Engagement Rate (n=24 Elementary Classrooms)

	STAR Teaching and Learning Components																
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Achievement Gain Index	.23	.36	.03	.32	.48*	.32	-.02	.17	.08	.04	.36	.26	.06	.13	.42*	.54**	.31
Engagement Rate																	
Quantity of Engagement	.51**	.48*	.66**	.58**	.66**	.19	.34	.57**	.60**	.31	.61**	.41*	.71**	.37	.43*	.21	.06
Quality of Engagement																	
High	.08	.32	.09	.46*	.25	.16	-.18	.27	.32	.09	.30	.08	.05	.32	.26	.54**	.09
Mid	-.06	-.33	-.02	-.28	-.14	.01	.15	-.11	.27	-.08	-.05	.04	-.07	-.21	-.16	-.[illegible]	-.04
Low	-.31	-.28	-.31	-.49**	-.52**	-.41*	-.12	-.55**	-.40*	-.14	-.62**	-.36	-.28	-.45*	-.33	-.68**	-.22

*p<.05
**p<.01

BEST COPY AVAILABLE

TABLE 9

Summary of Intercorrelations Between STAR Teaching and Learning Components and Subscales of the Classroom Learning Environment Scale (n=24 Middle and Secondary Schools)

	STAR Teaching and Learning Components																
Classroom Learning Environment Scale	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Autonomy	.06	-.04	.08	.12	.07	.05	.03	-.08	.32	.54**	-.01	.13	.41	.30	.12	.18	.04
Prior Knowledge	-.15	-.03	-.19	-.30	-.07	-.07	.15	.28	.52*	.49*	.42	.26	.20	.37	.26	.09	-.33
Collaboration	-.15	.15	.06	-.30	-.03	.28	.31	.07	.27	.32	.10	.19	.04	.08	-.04	.07	-.26
Reflection	-.20	.07	-.19	-.12	-.09	.18	.26	.32	.56**	.31	.60**	.39	.28	.33	.32	.13	-.30

*p<.05
**p<.01

TABLE 10

Summary of Intercorrelations Between STAR Teaching and Learning Components, Achievement Gain Index and Class Engagement Rate (n=20 Middle and Secondary Schools)

	STAR Teaching and Learning Components															
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Achievement Gain Index	-.26	.04	-.03	.01	-.15	.19	.13	.24	.08	.13	.17	.21	-.02	.11	.15	.07	.06
Engagement Rate																	
Quantity of Engagement	.63**	.56**	.70**	.64**	.63**	.25	.52*	.14	.30	.36	.18	-.02	.46*	.28	.35	.32	.43
Quality of Engagement																	
High	.47*	.52*	.36	.26	.28	.11	.46	-.02	.15	.24	.08	-.35	.08	-.11	.09	.13	-.01
Mid	-.17	.10	-.06	.13	-.17	.19	.14	.12	-.14	-.40	.18	.29	-.09	-.11	.10	-.08	-.19
Low	-.38	-.49*	-.01	-.15	-.32	-.30	-.62**	.07	-.53*	-.52*	-.36	.13	-.40	-.21	-.40	-.12	-.11

*p<.05
**p<.01

BEST COPY AVAILABLE

Table 11

Summary of Intercorrelations Between STAR Teaching and
Learning Components Achievement Gain Index and Class Engagement Rates for All Classrooms (n=66)

	STAR Teaching and Learning Components																
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Achievement Gain Index (n=43)	11	.36*	.09	.37**	.39**	.43**	.20	.31*	.27	.33*	.38**	.34*	.27	.27	.41**	.34*	.25
Class Engagement Rate																	
Quantity of Engagement	.48**	.54**	.69**	.45**	.55**	.28*	.10	.42**	.34**	.23	.46**	.19	.53**	.32**	.37**	.34**	.23
Quality of Engagement																	
Hi	.30*	.19	.15	.16	.10	-.01	.04	.01	.10	.01	.09	.23	.02	-.03	.05	.29*	-.01
Mid	-.13	.07	.00	-.03	-.06	.15	.15	.07	-.07	.00	.09	.26*	.00	.03	.04	-.23	.03
Lo	-.30*	-.42**	-.23	-.24	-.30*	-.33**	-.31**	-.25*	-.39**	-.33**	.55**	-.14	-.26*	-.35**	-.34**	-.26*	-.03

**p<.01
*p<.05

BEST COPY AVAILABLE

TABLE 12

Summary of Intercorrelations Between Subscales of the My Class Inventory (MCI), Achievement Gain Index and Class Engagement Rate Indices for Elementary (n=24) Classrooms

MCI Subscales	Class Engagement Rate				Achievement Gain Index	
	Quantity	Quality			Elementary	
		Hi	Mid	Lo	By Student (n=502)	By Class (n=24)
Cohesiveness	.48*	.25	-.24	-.42*	.11**	.44*
Friction	-.01	.01	.15	.14	-.23**	-.52**
Difficulty	.08	.52**	-.50**	.10	.08	.20
Satisfaction	.43*	.05	.04	-.39	.15**	.40*
Competitiveness	.33	.27	-.09	-.05	.04	.07
Achievement Gain Index	.15	.29	-.26	-.49**	...	...

**p<.01

BEST COPY AVAILABLE

TABLE 13

Summary of Intercorrelations Between Subscales of the Classroom Learning Environment Scale, Achievement Gain Index and Class Engagement Rate Indices for Middle and Secondary (n=20) Classrooms

CLES Subscales	Class Engagement Rate				Achievement Gain Index	
	Quantity	Quality			Middle/Secondary	
		Hi	Mid	Lo	By Student (n=280)	By Class (n=19)
Autonomy	.21	-.03	-.10	-.24	.04	.16
Prior Knowledge	-.27	-.17	-.07	-.28	.05	.31
Collaboration	.05	.23	.16	-.47*	.02	.31
Reflection	-.24	-.05	.23	-.42	.02	.32
Achievement Gain Index	-.02	-.17	.35	-.12	---	---

*p<.05

TABLE 14
Summary of Percentages of Maximum Possible Scores for Each STAR Teaching and Learning Component for "Superior", "Random" and "Comparison" Teachers Summed Over All Possible Assessment Decisions

STAR Teaching and Learning Components		Teacher Groups: "Superior" (n=34) Max.[b] Possible	"Superior" %Max.[c] Possible	"Random" (n=35) Max. Possible	"Random" %Max. Possible	"Comparison" (n=19) Max. Possible	"Comparison" %Max. Possible
II.A.	Time (8)[a]	1632	86	1680	80	912	75
II.B.	Classroom Routines(4)	816	89	840	84	456	77
II.C.	Student Engagement(1)[d]	204	68	210	60	114	60
II.D.	Managing Task-Related Behavior(6)	1224	78	1224	71	684	57
II.E.	Monitoring and Maintaining Student Behavior (9)	1836	79	1890	70	1026	63
III.A.	Psychosocial Learning Environment(12)	2248	94	2520	77	1368	69
III.B.	Physical Learning Environment (4)	816	94	840	88	456	82
IV.A.	Lesson and Activities Initiation(10)	2040	57	2100	47	1140	39
IV.B.	Teaching Methods and Learning Tasks(6)	1224	80	1260	72	684	62
IV.C.	Aids and Materials (8)	1632	87	1682	76	912	69
IV.D.	Content Accuracy and Emphasis(7)	1428	69	1470	64	798	53
IV.E.	Thinking Skills(11)	2244	46	2310	32	1252	22
IV.F.	Clarification(5)	1020	81	1050	75	570	59
IV.G.	Pace(3)	612	75	630	69	342	65
IV.H.	Monitoring and Informal Assessment(6)	1224	71	1260	62	684	49
IV.I.	Feedback(4)	816	59	840	49	456	43
IV.J.	Oral and Written Communication(4)	816	98	840	94	456	93

[a]Number of assessment indicators comprising component
[b]Maximum possible score = # of indicators x # of teachers x 6 assessments
[c]% of Max. possible = percentage of maximum possible decisions judged as "Acceptable"
[d]Maximum possible and % Max. Possible represent observed rates at or exceeding 90%

BEST COPY AVAILABLE

Table 15

Summary of One-Way ANOVA Results and Post Hoc Comparisons of Three Teacher Groups
(1 = "Superior"; 2 = "Random"; 3 = "Comparison") for Each STAR Teaching and Learning Component

STAR Teaching and Learning Component			F	p	Sheffe' Comparisons Significant at p<.05
II.	A.	Time (8)*	6.20	.0031	1 > 3
	B.	Classroom Routines (4)	4.01	.0216	1 > 3
	C.	Student Engagement (1)	0.67	.5149	—
	D.	Managing Task-Related Behavior (6)	4.42	.0149	1 > 3
	E.	Monitoring and Maintaining Student Behavior (9)	4.14	.0192	1 > 3
III.	A.	Psychosocial Learning Environment (12)	10.15	.0001	1 > 3, 2 > 3
	B.	Physical Learning Environment (4)	10.64	.0001	1 > 3, 2 > 3
IV.	A.	Lesson/Activities Initiation (10)	7.60	.0009	1 > 3
	B.	Teaching Methods and Learning Tasks (6)	6.63	.0021	1 > 3, 2 >3
	C.	Aids and Materials (8)	6.12	.0033	1 > 3
	D.	Content Accuracy and Emphasis (7)	8.61	.0004	1 > 3, 2 > 3

Table (continued)

STAR Teaching and Learning Component		F	p	Sheffe' Comparisons Significant at p<.05
E.	Thinking Skills (11)	11.06	.0001	1 > 2, 1 > 3
F.	Clarification (5)	9.22	.0002	1 > 3, 2 > 3
G.	Pace (3)	1.94	.1493	—
H.	Monitoring and Informal Assessment (6)	6.95	.0016	1 > 3, 2 > 3
I.	Feedback (4)	3.40	.0379	1 > 3
J.	Oral and Written Communication (4)	1.63	.2027	—

* Number of assessment indicators comprising component

Table 16

Comparison of Predicted Percentages of "Superior", "Random" and "Comparison" Teachers Below Recommended Standard for Each STAR Teaching and Learning Component

STAR Teaching and Learning Component			Performance Standard	Teacher Groups "Superior" (n=34)	"Random" (n=35)	"Comparison" (n=19)
II.	A.	Time (48) [a]	36 (75)[b]	8.8[c]	14.7	31.6
	B.	Classroom Routines (24)	18 (76)	5.9	17.6	31.6
	C.	Student Engagement___[d]	___	___	___	___
	D.	Managing Task-Related Behavior (36)	25 (70)	26.5	35.3	52.6
	E.	Monitoring and Maintaining Student Behavior (54)	38 (70)	23.5	44.1	47.4
III.	A.	Psychosocial Learning Environment (72)	55 (77)	14.7	32.4	63.2
	B.	Physical Learning Environment (24)	20 (83)	8.8	17.6	57.9
IV.	A.	Lesson/Activities Initiation (60)	43 (71)	79.4	88.2	100.0
	B.	Teaching Methods and Learning Tasks (36)	27 (74)	23.5	35.3	68.4
	C.	Aids and Materials (48)	36 (75)	14.7	26.5	57.9

Table (continued)

STAR Teaching and Learning Component		Performance Standard	Teacher Groups "Superior" (n=34)	"Random" (n=35)	"Comparison" (n=19)
D.	Content Accuracy and Emphasis (42)	32 (75)	70.6	73.5	100.0
E.	Thinking Skills (66)	44 (67)	85.3	94.1	100.0
F.	Clarification (30)	23 (75)	32.4	38.2	73.7
G.	Pace (18)	12 (74)	29.4	29.4	52.6
H.	Monitoring and Informal Assessment (36)	27 (75)	44.1	61.8	84.2
I.	Feedback (24)	18 (74)	82.4	79.4	84.2
J.	Oral and Written Communication (24)	20 (87)	2.9	5.9	10.5

[a] Maximum Possible Score for Component (3 assessments x 2 occasions)

[b] Percentage of Maximum Possible Score Recommended as a Performance Standard

[c] Predicted Percentage of Teachers Below Recommended Performance Standard

[d] Student Engagement Index is Not Recommended for Use for Certification

Table (continued)

STAR Teaching and Learning Component		Performance Standard	Teacher Groups "Superior" (n=34)	"Random" (n=35)	"Comparison" (n=19)
D.	Content Accuracy and Emphasis (42)	32 (75)	70.6	73.5	100.0
E.	Thinking Skills (66)	44 (67)	85.3	94.1	100.0
F.	Clarification (30)	23 (75)	32.4	38.2	73.7
G.	Pace (18)	12 (74)	29.4	29.4	52.6
H.	Monitoring and Informal Assessment (36)	27 (75)	44.1	61.8	84.2
I.	Feedback (24)	18 (74)	82.4	79.4	84.2
J.	Oral and Written Communication (24)	20 (87)	2.9	5.9	10.5

[a] Maximum Possible Score for Component (3 assessments x 2 occasions)

[b] Percentage of Maximum Possible Score Recommended as a Performance Standard

[c] Predicted Percentage of Teachers Below Recommended Performance Standard

[d] Student Engagement Index is Not Recommended for Use for Certification

83

Table 17

Generalizability Coefficients for the STAR Teaching/Learning Components

Teaching/ Learning Component	G-Coefficient: Principal and External Assessor	G-Coefficient Principal, External Assessor and Master Teacher
# 7 Time	.598	.643
# 8 Classroom Routines	.525	.577
#10 Managing Task-Related Behavior	.645	.700
#11 Monitoring/Maintaining Student Behavior	.723	.775
#12 Psychosocial Learning Environment	.726	.789
#13 Physical Learning Environment	.631	.695
#14 Lessons/Activities Initiation	.664	.722
#15 Teaching Methods	.577	.630
#16 Sequence/Pace	.521	.576
#17 Aids and Materials	.614	.682
#18 Content Accuracy/ Emphasis	.660	.728
#19 Thinking Skills	.732	.807
#20 Clarification	.447	.497
#21 Monitoring Learning Activities/Informal Assessment	.596	.651
#22 Feedback	.625	.691
#23 Oral/Written Communication	.130	.147

NOTE: Both models presented here simulate a three observer model. The second model adds the effect of the third observer (master teacher) to that of the first two observers (principal and external assessor).

TABLE 18

Generalizability Coefficients for the STAR Teaching and Learning Components[75]

Teaching and Learning Components		G-Coefficient Principal and External Assessor	G-Coefficient Principal, External Assessor and Master Teacher
PERFORMANCE DIMENSION II: CLASSROOM AND BEHAVIOR MANAGEMENT			
A.	Time	0.223	0.292
B.	Classroom Routines	0.441	0.540
D.	Managing Task-Related Behavior	0.595	0.683
E.	Monitoring and Maintaining Student Behavior	0.561	0.655
PERFORMANCE DIMENSION III: LEARNING ENVIRONMENT			
A.	Psychsocial	0.461	0.557
B.	Physical	0.30	0.391
PERFORMANCE DIMENSION IV: ENHANCEMENT OF LEARNING			
A.	Lesson and Activities Initiation	0.397	0.497
B.	Teaching Methods and Learning Tasks	0.616	0.702
C.	Aids and Materials	0.386	0.463
D.	Content Accuracy and Emphasis	0.383	0.483
E.	Thinking Skills	0.433	0.526
F.	Clarification	0.327	0.419
G.	Pace	0.268	0.355
H.	Monitoring Learning Tasks and Informal Assessment	0.560	0.647
I.	Feedback	0.370	0.462
J.	Oral and Written Communication	0.340	0.435

TABLE 19

Percentage of Acceptable Decisions for Teaching/Learning Components for each Dimension of the STAR for LSU GTAs and Louisiana Public School Teachers

TEACHING/LEARNING COMPONENTS		# of Indicators	%of Maximum* LSU GTAs (n=25)	%of Maximum* LA Teachers (n=6000)
PERFORMANCE DIMENSION II CLASSROOM BEHAVIOR AND MANAGEMENT				
A.	Time	6	68	72
B.	Classroom Routines	4	81	74
C.	Managing Task-Related Behavior	6	38	48
D.	Monintoring/Maintaining Student Behavior	6	50	54
PERFORMANCE DIMENSION III LEARNING ENVIRONMENT				
A.	Psychosocial	10	65	66
B.	Physical	3	67	88
PERFORMANCE DIMENSION IV ENHANCEMENT OF LEARNING				
A.	Lesson/Activities Initiation	8	22	34
B.	Teaching Methods and Learning Tasks	6	46	59
C.	Aids and Materials	6	47	62
D.	Content Accuracy/Emphasis	6	46	49
E.	Thinking Skills	11	12	22
F.	Clarification	4	53	54
G.	Monitoring Learning Tasks/Informal Assessment	6	17	43
H.	Feedback	4	24	33
I.	Oral/Written Communication	4	89	95

*Index computed by dividing actual obtained score by maximum possible score